

BAB I

PENDAHULUAN

1.1 Latar Belakang

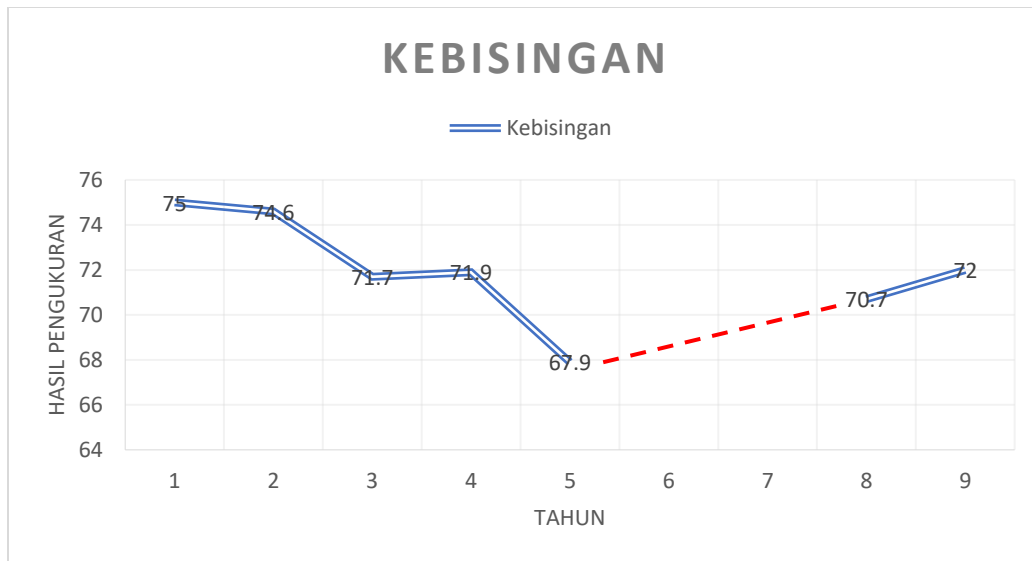
Saat ini telah memasuki era big data (Davenport, 2014), perkembangan big data tidak hanya mengenai jumlah/*volume* data yang begitu besar, melainkan karakteristik lain yang menggambarkan big data yaitu *velocity* dan juga *variability* data (hen, 2013). *Variability* data merupakan salah satu karakteristik *big data* yang menggambarkan bahwa data yang begitu besar dan banyak, lazimnya didapatkan dari beberapa sumber yang berbeda (George, 2014). Dalam *variability* data terdapat tiga level tingkatan jenis data yaitu terstruktur, semi-terstruktur, dan juga tidak terstruktur (Coronel, 2013). Data tidak terstruktur merupakan data yang sulit dianalisis karena mengandung beberapa kondisi data salah satunya yaitu data yang tidak lengkap dan juga data yang hilang (Samuel, McAbee, Landis, & Burke, 2016).

Badan Lingkungan Hidup merupakan suatu badan dibawah pemerintahan yang memiliki tugas utama membantu kepala daerah dalam merumuskan kebijakan dan melakukan koordinasi di bidang lingkungan hidup. Tugas tersebut sesuai dengan Peraturan Pemerintah Nomor 38 Tahun 2007 tentang pembagian urusan pemerintahan, antara pemerintah daerah, pemerintah provinsi dan pemerintah kabupaten atau kota. Selain itu, terdapat juga tugas khusus yang perlu untuk dilakukan oleh Badan Lingkungan Hidup di setiap provinsi salah satunya yaitu melakukan pengawasan dan pengendalian pengelolaan kualitas udara dan pengendalian pencemaran udara (Badan Lingkungan Hidup, 2015). Badan Lingkungan Hidup provinsi Daerah Istimewa Yogyakarta telah rutin melakukan tugas

tersebut dengan membuat laporan hasil pengukuran kualitas udara secara tahunan. Pengukuran mengenai kualitas udara ditentukan oleh beberapa parameter yang telah ditentukan antara lain kebisingan, kelembapan, suhu udara, arah angin, dan kecepatan angin, kadar Sulfur Dioksida (SO₂), konsentrasi debu PM.10, Timbal Hitam (Pb), Nitrogen Dioksida (NO₂), dan Hidrokarbon (HC).

Missing value adalah suatu kondisi dimana data tersebut tidak lengkap yang disebabkan oleh beberapa kondisi tertentu. Saat ini *missing value* merupakan kondisi yang umum terjadi pada permasalahan perkumpulan data (*dataset*) di berbagai bidang seperti hidrologi, *engineering*, *electricity marketing*, *economics* dan juga transportasi (Aydilek & Arslan, 2013). *Missing* akan memberikan dampak terhadap kualitas dari sebuah penelitian. Menurut Luengo, et al., (2012) suatu penelitian yang mengandung *missing value* akan mendapatkan dampak bahwa penelitian tersebut nilai efisiensinya akan berkurang, proses analisisnya jauh lebih kompleks, kompleks dalam pengaturannya, dan hasil yang didapatkan akan menimbulkan hasil yang bias jika dibandingkan dengan data yang tidak mengandung *missing value*.

Di beberapa negara besar, keberadaan data parameter kualitas udara tersebut digunakan untuk melakukan peramalan (*forecasting*), peramalan kualitas udara adalah cara yang efektif untuk melindungi kesehatan masyarakat dengan menyediakan peringatan dini terhadap polutan udara berbahaya (Yun Bai, 2016). Langkah inilah yang belum dilakukan oleh pihak Badan Lingkungan Hidup provinsi Daerah Istimewa Yogyakarta. Belum dilakukan hal ini dikarenakan adanya beberapa data yang tidak lengkap ataupun hilang (*missing value*). Berdasarkan hasil observasi langsung dapat diketahui bahwa di beberapa tahun serta titik pengamatan menunjukkan data yang tidak lengkap. Berikut merupakan salah satu contoh data yang tidak lengkap/hilang (*missing value*) disuatu titik pengukuran yang dilakukan di Depan Toko Besi Selatan Dongkelan Jalan Bantul, Yogyakarta :



Gambar 1. 1 *Missing Value* pada titik pengamatan

Sumber : Badan Lingkungan Hidup, 2015

Missing value yang terdapat di beberapa titik dan tahun pengamatan tentunya mengganggu adanya proses pembukuan serta pembangunan *database* yang sedang dicanangkan oleh pihak Badan Lingkungan Hidup provinsi Daerah Istimewa Yogyakarta. Selain itu, proses peramalan kualitas udara untuk tahun mendatang sebagai langkah antisipasi dini terhadap polutan yang berbahaya juga akan terhambat karena nantinya hasil yang didapatkan akan menunjukkan nilai yang bias. Pentingnya menangani data yang hilang terletak dalam menemukan beberapa sifat tersembunyi dari data untuk lebih menghindari kesalahan dalam statistik dan peramalan data (Qin, Liu, Wang, & Sun, 2014).

Interpolasi merupakan jenis metode untuk membangun titik data baru melalui himpunan data yang ada (Tan, 2002). Untuk itu dalam upaya mengembalikan data yang hilang tersebut akan dilakukan interpolasi. Terdapat beberapa bentuk ataupun model interpolasi yang dapat digunakan dalam proses mengembalikan data yang hilang (*missing value*) tersebut. Dengan membentuk suatu model persamaan interpolasi maka mengestimasi nilai yang hilang (*missing value*) dapat dilakukan. Model persamaan akan membentuk suatu persamaan tertentu berdasarkan kondisi data yang diketahui nilainya sebelumnya dan dapat digunakan untuk mengestimasi nilai yang hilang tersebut. Untuk itu, dalam pembentukan model persamaan interpolasi perlu adanya suatu pembentukan persamaan interpolasi yang tepat untuk menghasilkan estimasi data yang tepat.

Penelitian sebelumnya telah dilakukan terkait dengan penggunaan interpolasi untuk mengembalikan data yang hilang (*missing value*) terkait dengan data kualitas udara. Penelitian yang dilakukan oleh Junninenena, et al., (2004) telah membandingkan antara dua metode yang berbeda yaitu antara univariate method dengan multivariate method. Metode univariat menggunakan beberapa metode yaitu *linier interpolation*, *spline interpolation*, dan *nearest neighbour interpolation*. Sementara itu, pada metode multivariat menggunakan regresi, multivariat *nearest neighbour*, *self-organizing map* dan *multi perceptron layer*. Berdasarkan hasil yang didapatkan menunjukkan bahwa pada hasil metode univariat yang dilakukan tergantung dengan lebar jarak kesenjangan pada data yang hilang dan membutuhkan metode lain untuk memperkirakan lebar jarak kesenjangan agar didapatkan hasil yang lebih baik. Sementara itu, pada metode multivariat menunjukkan hasil yang lebih baik jika melihat nilai akurasi yang didapatkan.

Penelitian lain telah dilakukan oleh Galan, et al., (2016) yang membuktikan bahwa metode *Genetic Algorithm* juga mampu digunakan sebagai metode untuk mengestimasi data yang hilang. Dalam penelitian data yang hilang yang akan dicari yaitu data kuesioner yang tidak lengkap diisi oleh penulisnya. Hal ini membuktikan bahwa tidak hanya data yang berupa data *time series* juga dapat diestimasi nilainya. Hasil yang didapatkan menunjukkan bahwa hasil mengestimasi data yang hilang dapat dilakukan dengan melibatkan dua hingga empat kriteria untuk menghasilkan satu nilai yang dicari. Selain itu, dijelaskan juga mengenai hasil metode *Genetic Algorithm* yang lebih efektif jika dibandingkan dengan metode MICE yang sebelumnya biasa digunakan dalam mengestimasi *missing value* pada konteks kuesioner.

Penelitian yang dilakukan oleh Paul, et al., (2017) menjelaskan bahwa dalam mengestimasi data yang hilang dapat juga dilakukan menggunakan metode klasifikasi. Kasus yang diangkat yaitu mengenai nilai gen yang hilang. Dalam penelitian ini, metode yang digunakan adalah *Fuzzy Importance Factor (FIF)* dengan menggunakan algoritma PSM. Dan hasil yang didapatkan menunjukkan bahwa metode ini menunjukkan hasil yang baik dalam beberapa variabel antara lain tingkat akurasi, nilai *error* yang kecil, dan juga sensitivitas yang baik.

Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya dapat membuktikan bahwa penelitian mengenai mengestimasi nilai yang hilang (*missing value*) perlu untuk dilakukan.

Dalam penelitian ini metode yang akan digunakan adalah dua metode yakni regresi linier dan *curve fitting method*. Metode yang akan digunakan disesuaikan dengan kondisi data serta dependensi antar variabel yang diketahui. Data yang akan digunakan dalam penelitian ini yakni sulfur dioksida (SO₂), nitrogen dioksida (NO₂), hidrokarbon (HC), timah hitam (Pb), dan debu diameter (PM₁₀) yang dipengaruhi oleh variabel meteorologi atau aspek fisika yang meliputi suhu udara, kelembapan, dan kecepatan angin di Provinsi Daerah Istimewa Yogyakarta karena studi terdahulu menunjukkan bahwa variasi aspek meteorologi, seperti arah angin, kecepatan angin, dan kelembapan relatif dapat mempengaruhi variasi pada kualitas polutan udara (Elminir, 2005), selain itu dapat juga disebabkan karena sumber polutan (Altunkaynak, Ozger, & Cakmakci, 2005). Nantinya, kedua metode tersebut digunakan untuk membuat suatu model persamaan interpolasi yang tepat dalam mengestimasi nilai yang hilang (*missing value*) pada beberapa variabel tersebut. Dari hasil interpolasi kedua metode tersebut dapat diketahui hasil yang terbaik dengan mengetahui nilai *error* yang dihasilkan dengan melihat nilai MAPE (*Mean Absolute Percentage Error*) untuk setiap variabel yang telah diestimasi nilainya. Hasil interpolasi yang dipilih akan disesuaikan dengan nilai *error* yang dihasilkan yang menunjukkan nilai terkecil.

Penelitian ini adalah penelitian yang masih dalam lingkup keilmuan Teknik Industri. Penelitian mengenai mengestimasi suatu data yang hilang (*missing value*) termasuk salah satu pengetahuan mengenai *data processing*. *Data processing* adalah suatu kajian untuk membantu ataupun menunjang dalam basis efisiensi serta produktifitas dalam kajian Teknik Industry (Salvendy, 1992).

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana metode yang tepat untuk membentuk model persamaan dalam melakukan interpolasi data yang hilang (*missing value*)?
2. Bagaimana hasil interpolasi pada data yang hilang (*missing value*) di beberapa tahun yang digunakan?

3. Bagaimana pola kecenderungan nilai *error* masing-masing variabel dari model persamaan yang terbentuk?

1.3 Ruang Lingkup Penelitian

Ruang Lingkup dalam penelitian ini adalah

1. Data mengenai kualitas udara ambien hanya pada titik – titik tertentu yang terdapat pada laporan tahunan oleh Badan Lingkungan Hidup Provinsi DIY
2. Data yang digunakan hanya pada tahun 2009 – 2015.

1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah :

1. Mendapatkan metode yang tepat untuk membentuk model persamaan dalam melakukan prediksi data yang hilang (*missing value*).
2. Mengetahui hasil prediksi pada data yang hilang (*missing value*) di beberapa tahun yang digunakan.
3. Mengetahui pola kecenderungan nilai *error* masing – masing variabel dari model persamaan yang terbentuk.

1.5 Manfaat Penelitian

Manfaat dalam penelitian ini adalah

1. Data hasil interpolasi untuk data yang hilang (*missing value*) dapat digunakan untuk penelitian berikutnya.

2. Model persamaan yang terbentuk dapat digunakan sebagai referensi untuk data yang hilang (*missing value*) pada kasus yang lain.

1.6 Sistematika Penulisan

Dalam penelitian ini menggunakan kaidah penulisan ilmiah dengan sistematika penulisan sebagai berikut.

BAB I PENDAHULUAN

Pada bagian ini memuat latar belakang, rumusan masalah, ruang lingkup penelitian, tujuan penelitian, manfaat penelitian dan sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Bab ini memuat kajian literatur deduktif dan induktif berupa teori-teori dan penelitian-penelitian terdahulu untuk memposisikan penelitian ini dan menjadi acuan dalam menyelesaikan permasalahan yang ada pada penelitian ini

BAB III METODOLOGI PENELITIAN

Bagian ini memuat obyek penelitian, jenis data dan metode yang digunakan, serta alur penelitian yang dilakukan dalam penelitian secara ringkas dan jelas.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Bagian ini berisi tentang uraian pengumpulan dan proses pengolahan data dengan prosedur tertentu. Hasil pengolahan data ditampilkan dalam bentuk tabel dan grafik. Bab ini menjadi acuan untuk melakukan pembahasan hasil pada bab selanjutnya.

BAB V PEMBAHASAN

Bab ini berisi pembahasan secara detail dan kritis mengenai hasil pengolahan data pada bab sebelumnya dan belum dipaparkan di bab sebelumnya. Pembahasan dilakukan untuk menjawab rumusan masalah yang ada dan memberikan rekomendasi berdasarkan hasil pengolahan data.

BAB VI KESIMPULAN DAN SARAN

Bagian ini berisi tentang kesimpulan yang menjabarkan hasil penelitian secara singkat dan jelas serta rekomendasi mengenai pengembangan yang dapat dilakukan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

LAMPIRAN