

**IMPLEMENTASI REGRESI LINEAR UNTUK MEMPREDIKSI
LAMA WAKTU PENGIRIMAN *CATERING* KEPADA
KONSUMEN STUDI KASUS HOME CATERING
MALANG**

SKRIPSI

**OLEH:
FERY FENGKI
NIM. 11650069**



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2018**

HALAMAN PENGAJUAN

**IMPLEMENTASI REGRESI LINEAR UNTUK MEMPREDIKSI LAMA
WAKTU PENGIRIMAN CATERING KEPADA KONSUMEN
STUDI KASUS HOME CATERING MALANG**

SKRIPSI

**Diajukan Kepada :
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**OLEH:
FERY FENGKI
NIM. 11650069**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2018**

HALAMAN PERSETUJUAN

**IMPLEMENTASI REGRESI LINEAR UNTUK MEMPREDIKSI LAMA
WAKTU PENGIRIMAN CATERING KEPADA KONSUMEN
STUDI KASUS HOME CATERING MALANG**

SKRIPSI

Oleh:

FERY FENGKI


NIM. 11650069


Telah Diperiksa dan Disetujui untuk Diuji

Tanggal: 4 Juli 2018

Dosen Pembimbing I

Dosen Pembimbing II


Dr. Cahyo Crysdiان
NIP. 19740424 200901 1 008



M. Imamuddin, Lc., MA
NIP. 19740602 200901 1 010

Mengetahui,

Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi

Universitas Maulana Malik Ibrahim Malang




Dr. Cahyo Crysdiان
NIP. 19740424 200901 1 008

HALAMAN PENGESAHAN

**IMPLEMENTASI REGRESI LINEAR UNTUK MEMPREDIKSI LAMA
WAKTU PENGIRIMAN CATERING KEPADA KONSUMEN
STUDI KASUS HOME CATERING MALANG**

SKRIPSI

Oleh :
FERY FENGKI
NIM. 11650069

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal : Juli 2018

Susunan Dewan Penguji




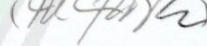
Penguji Utama : Fresy Nugroho, M.T
NIP.19710722 201101 1 001

Ketua Penguji : A'la Syauqi, M.Kom
NIP. 19771201 200801 1 007

Sekretaris Penguji : Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

Anggota Penguji : M. Imamuddin, Lc., MA
NIP.19740602 200901 1 010


Tanda Tangan

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi

Universitas Islam Maulana Malik Ibrahim Malang




Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

**HALAMAN PERNYATAAN
ORISINALITAS PENELITIAN**

Saya yang bertanda tangan dibawah ini :

Nama : Fery Fengki

NIM : 1650069

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Judul : **Implementasi Regresi Linear untuk Memprediksi Lama Waktu Pengiriman Catering kepada Konsumen Studi Kasus Home Catering Malang**

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini tidak terdapat unsur-unsur penjiplakan karya penelitian atau karya ilmiah yang pernah dilakukan atau dibuat oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka. Apabila ternyata hasil penelitian ini terbukti terdapat unsur-unsur penjiplakan, maka saya bersedia untuk mempertanggungjawabkan, serta diproses sesuai peraturan yang berlaku.

Malang, 8 Juni 2018
Yang membuat pernyataan



Fery Fengki
NIM. 11650069

MOTTO

“You can't go back and change the beginning, but you can start where you are and change the ending”

C.S. Lewis



HALAMAN PERSEMBAHAN

Alhamdulillah robbil'alamiin, puji syukur kehadirat Allah SWT yang telah memberikan nikmat, kekuatan, dan semangat tiada tara untuk menyelesaikan kewajiban saya menuntut ilmu di kampus ini. Dengan segenap rasa syukur yang mendalam, skripsi ini saya persembahkan untuk :

Ibuku Drusni dan Bapakku Samhadi

atas berkah doa restu serta dukungan moril dan materiil
dan semua hal yang tak ternilai

Adik, Saudara, dan Teman Dekat

atas doa dan canda guranya yang membuat hiruk pikuk di rumah,
di desa saat pulang ke kampung halaman.

Teman-teman TI 2011, Saudara dan Kerabatku

yang telah memberikan support, dukungan dan doa. Saya ucapkan banyak
terima kasih kepada kalian semua.

Teman-Teman yang Berkontribusi

yang telah memberikan *support*, dukungan dan doa. Saya ucapkan banyak
terima kasih kepada kalian semua. Berbagai dukungan secara langsung maupun tidak
langsung. Untuk teman satu angkatan yang telah *graduate* duluan tapi masih sering
memberi *support*, terima kasih atas kepeduliannya *KF

***Kepada Seluruh Dosen dan Pihak yang Berjasa, dalam
Menyelesaikan Skripsi ini secara Langsung maupun
Tidak Langsung***

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Segala puji bagi Allah SWT atas Rahmat dan Hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini. Sholawat dan Salam tetap tercurahkan Nabi Muhammad SAW, sang pemberi syafaat kelak di hari akhir, beserta seluruh keluarga, sahabat, dan para pengikutnya.

Penelitian skripsi yang berjudul *“Implementasi Regresi Linear untuk Memprediksi Lama Waktu Pengiriman Catering kepada Konsumen Studi Kasus Home Catering Malang”* ini ditulis untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Strata Satu (S1) Fakultas Sains dan Teknologi Universitas Maulana Malik Ibrahim Malang. Dengan selesainya skripsi ini penulis mengucapkan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Cahyo Crysdiyan, selaku Ketua Jurusan dan dosen pembimbing I yang telah meluangkan waktu untuk membimbing, memotivasi, mengarahkan dan memberi masukan dalam pengerjaan skripsi ini hingga akhir dan telah bersedia meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, berbagai pengalaman, arahan, nasihat, motivasi dan pengarahan dalam pembangunan program hingga penyusunan skripsi ini.
3. Bapak M. Imamuddin, Lc., MA, selaku dosen pembimbing 2 yang memberi masukan, serta pengarahan dalam penyusunan laporan skripsi ini.

4. Ibuku dan Bapakku dan adik-adikku serta keluarga besar tercinta yang selalu memberi dukungan yang tak terhingga serta doa yang senantiasa mengiringi setiap langkah.
5. Segenap civitas akademika Fakultas Saintek, Universitas Islam Negeri Maulana Malik Ibrahim Malang terutama seluruh dosen, terima kasih atas segala ilmu dan bimbingannya.
6. Seluruh teman-teman Informatika terutama angkatan 2011, terima kasih atas pertemanan, dukungan dan kebersamaan yang luar biasa, hingga saya bisa mencapai gelar ini.

Harapan penulis semoga semua amal kebaikan dan jasa-jasa dari semua pihak yang telah membantu hingga skripsi ini selesai diterima oleh Allah SWT, serta mendapatkan balasan yang lebih baik dan berlipat ganda.

Penulis juga menyadari bahwa skripsi ini masih jauh dari kesempurnaan yang disebabkan keterbatasan penulis, semoga karya ini bermanfaat dan menambah ilmu pengetahuan bagi kita semua, Aamiin.

Malang, 4 Juli 2018

Penulis

Fery Fengki

DAFTAR ISI

HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN ORISINALITAS PENELITIAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
ABSTRAK.....	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
BAB 2 KAJIAN PUSTAKA.....	6
2.1 Landasan Teori	6
2.1.1 Prediksi.....	6
2.1.2 <i>Data Mining</i>	10
2.1.3 Metode Regresi Linear	15
2.1.4 Regresi Linear Sederhana	16
2.1.5 Regresi Linear Berganda.....	17
2.1.6 Asumsi Klasik dalam Regresi Linear.....	18
2.1.6.1 Multikolinearitas	18
2.1.6.2 Heteroskedastisitas	23
2.1.6.3 Autokorelasi	26
2.1.6.4 Normalitas	27

2.1.7	Akurasi Prediksi	27
2.2	Profil Home Catering Malang	29
2.3	Penelitian Terkait	31
BAB 3	METODE PENELITIAN	33
3.1	Sumber Data	33
3.2	Objek Penelitian	34
3.3	Analisis dan Desain Sistem	34
3.3.1	Desain Sistem.....	34
3.3.1.1	Data Pengiriman.....	35
3.3.1.2	Praproses Data.....	35
3.3.1.3	<i>Training</i> Data dengan Metode Regresi Linear.....	38
3.3.1.4	Persamaan Regresi Linear.....	41
3.3.1.5	<i>Testing</i> Data	41
3.3.1.6	Hasil Perbandingan Y awal dengan Y hasil <i>training</i>	42
BAB 4	HASIL DAN PEMBAHASAN	43
4.1	Implementasi Antarmuka/ <i>Interface</i>	43
4.1.1	Tampilan <i>Load Data</i>	44
4.1.2	Tampilan <i>Data Training</i>	44
4.1.3	Tampilan <i>Input</i> untuk Memprediksi Lama Waktu Pengiriman	45
4.2	Implementasi Sistem	46
4.2.1	Proses <i>Load Data</i>	47
4.2.2	Proses <i>Training Data</i>	48
4.2.3	Prediksi Lama Waktu Pengiriman	51
4.3	Uji Asumsi Statistik.....	54
4.4	Uji Akurasi Hasil Prediksi.....	60
4.5	Integrasi dalam Islam	68
BAB 5	KESIMPULAN	70
5.1	Kesimpulan	70
5.2	Saran.....	70
DAFTAR PUSTAKA	71

DAFTAR GAMBAR

Gambar 2.1 Gambaran proses prediksi	10
Gambar 2.2 Disiplin ilmu dalam data mining	12
Gambar 2.3 Proses <i>data mining</i>	15
Gambar 2.4 Garis regresi Y karena pengaruh X, persamaan	17
Gambar 2.5 Contoh Homoskedastisitas	24
Gambar 2.6 Contoh Heteroskedastisitas.....	24
Gambar 2.7 Logo Home Catering Malang	30
Gambar 3.1 Desain sistem.....	35
Gambar 3.2 Proses Training	39
Gambar 4.1 Halaman Utama Sebelum Data Diproses	43
Gambar 4.2 Tampilan Load Data	44
Gambar 4.3 Tampilan Data Training.....	45
Gambar 4.4 Tampilan prediksi lama waktu pengiriman	46
Gambar 4.5 Halaman antar muka EViews 7	55

DAFTAR TABEL

Tabel 3.1 Data sampel pengiriman harian.....	34
Tabel 3.2 Model data daerah 1 (Soekarno Hatta).....	35
Tabel 3.3 Model data daerah 2 (Dinoyo)	36
Tabel 3.4 Model data daerah 3 (Tlogomas)	36
Tabel 3.5 Model data daerah 4 (Sumbersari)	36
Tabel 3.6 Model data daerah 5 (BCT).....	37
Tabel 3.7 Nilai Konstanta dan koefisien variabel x	40
Tabel 3.8 Koefisien Regresi Tiap Daerah	41
Tabel 3.9 Perbandingan nilai Y awal dengan Y training.....	42
Tabel 4.1 Kode program proses load data.....	47
Tabel 4.2 Training data daerah 1 (Soekarno-Hatta)	49
Tabel 4.3 Daftar persamaan regresi yang terbentuk.....	50
Tabel 4.4 Proses Prediksi lama waktu pengiriman.....	52
Tabel 4.5 Pengujian untuk daerah Soekarno-Hatta	56
Tabel 4.6 Hasil uji asumsi klasik.....	57
Tabel 4.7 Uji asumsi klasik untuk masing-masing daerah	58
Tabel 4.8 Nilai koefisien determinasi pada setiap persamaan.....	59
Tabel 4.9 Pengujian dengan persamaan 1	61
Tabel 4.10 Pengujian dengan persamaan 2	62
Tabel 4.11 Pengujian dengan persamaan 3	63
Tabel 4.12 Pengujian dengan persamaan 4	65
Tabel 4.13 Pengujian dengan persamaan 5	66
Tabel 4.14 Perbandingan hasil uji akurasi.....	67

ABSTRAK

Fengki, Fery. 2018. *Implementasi Regresi Linear untuk Memprediksi Lama Waktu Pengiriman Catering kepada Konsumen Studi Kasus Home Catering Malang*. Skripsi. Teknik Informatika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) Dr. Cahyo Crysdiyan, (II) M. Imamuddin, Lc., MA

Kata Kunci: Regresi Linear, Prediksi, Data Mining, Catering, Pengiriman

Home Catering Malang adalah wirausaha bisnis rumahan yang bergerak di bidang industri makanan. Memiliki banyak pelanggan dan harus mengantarkan ratusan kotak makanan setiap harinya, tentunya harus bisa memberikan layanan positif yang memuaskan bagi pelanggannya untuk senantiasa menjaga kredibilitas bisnis. Cara agar tetap bisa menjaga kredibilitas salah satunya yaitu dengan mengantarkan makanan secara tepat waktu. Permasalahan kemudian muncul saat setiap hari sering terjadi fluktuasi jumlah pelanggan, perkembangan jumlah pelanggan sulit diprediksi seiring berjalannya waktu. Menghadapi kondisi seperti itu tentu harus cermat dalam mengatur waktu karena bisa berpengaruh terhadap kecepatan pengiriman. Realitanya, banyak pelanggan yang mulai mengeluhkan layanan pengiriman yang semakin lama semakin sulit untuk konsisten, pelanggan banyak yang komplain tentang jam pengiriman yang melewati batas waktu maksimal. Oleh karenanya, dibutuhkan sebuah sistem yang dapat memprediksi berapa lama waktu yang dibutuhkan untuk menyelesaikan proses pengiriman kepada semua pelanggan berdasarkan faktor-faktor yang dinilai berpengaruh seperti jumlah pelanggan dan total jarak tempuh, serta lokasi dimana pelanggan berada. Dengan memanfaatkan data history pengiriman harian, sistem ini dapat melakukan analisa menggunakan metode regresi linear. Sehingga dapat menghasilkan persamaan linear yang dapat digunakan sebagai formula untuk memprediksi total waktu yang dibutuhkan untuk menyelesaikan proses pengiriman catering.

ABSTRACT

Fengki, Fery. 2018. *Linear Regression Implementation to Predict the Long Time Delivery Catering to Consumer Case Study Home Catering Malang. Essay. Technical Information*. Faculty of Science and Technology. State Islamic University Maulana Malik Ibrahim Malang.

Counselor: (I) Dr. Cahyo Crysdiان, (II) M. Imamuddin, Lc., MA

Keywords: Linear Regression, Prediction, Data Mining, Catering, Delivery

Home Catering Malang is a home business entrepreneur engaged in the food industry. Having many customers and having to deliver hundreds of boxes of food every day, it must be able to provide a satisfactory positive service for its customers to always maintain business credibility. One way to maintain credibility is one of them is to deliver food in a timely manner. Problems then arise when every day frequent fluctuations in the number of customers, the increase in the number of customers is difficult to predict over time. Facing such conditions would have to be careful in because it could have an effect on the speed of delivery. In reality, many customers are starting to complain about the increasingly difficult delivery service to be consistent, many customers complain about delivery hours that exceed the maximum deadline. Therefore, it takes a system that can predict how long it will take to complete the delivery process to all customers based on factors considered to be influential such as the number of customers and the total distance traveled, as well as the location where the customer is located. By utilizing daily delivery data history, this system can perform analysis using linear regression method. So it can produce a linear equation that can be used as a formula to predict the total time required to complete the catering delivery process

ملخص

Fery, Fengki . 2018. تنفيذ الانحدار الخطي للتنبؤ بالوقت للتسليم لتقديم الطعام إلى المستهلك دراسة المحلة *Home Catering Malang*. هندسة المعلوماتية. كلية العلوم والتكنولوجيا. جامعة الحكومية الإسلامية مولانا

مالك إبراهيم مالانج

المستشار: (الأول) Dr. Cahyo Crys dian (الثاني) M. Imamuddin, Lc. MA

الكلمات الأساسية : الانحدار الخطي ، التنبؤ ، استخراج البيانات ، تقديم الطعام ، التسليم

Home Catering مالانج هي شركة أعمال تجارية منزلية تعمل في مجال الصناعات الغذائية. لديها العديد من العملاء ويجب تقديم مئات من صناديق الطعام كل يوم، ويجب أن تكون قادرة على تقديم خدمة مرضية لعملائها إيجابي دائما الحفاظ على مصداقية العمل. فطريقة واحدة للحفاظ على المصداقية هي وتوصيل الطعام في الوقت المناسب. ثم تنشأ المشكلة عندما تزداد كل يوم تقلبات متكررة في عدد من العملاء، وعدد متزايد من العملاء من الصعب التنبؤ على مر الزمن. إن مواجهة مثل هذه الظروف يجب أن تكون حذرة لأنها قد يكون لها تأثير على سرعة التسليم. في الواقع، بدأ العديد من العملاء يشكون من تقديم الخدمات التي يصعب على نحو متزايد أن تكون متسقة، اشتكى العديد من العملاء حول تسليم الماضي ساعة الحد الأقصى الوقت. لذلك، نحن بحاجة إلى نظام يمكن التنبؤ كم من الوقت يستغرق لإتمام عملية التسليم لجميع العملاء على أساس عوامل تعتبر مؤثرة حيث بلغ عدد العملاء والمسافة الإجمالية، وكذلك المكان الذي يقع العميل. من خلال استخدام محفوظات بيانات التسليم اليومية ، يمكن لهذا النظام إجراء التحليل باستخدام طريقة الانحدار الخطي. وذلك لإنتاج معادلة الخطية التي يمكن استخدامها بوصفها صيغة للتنبؤ الوقت الإجمالي المطلوب لإتمام عملية التسليم الطعام.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan internet mendorong dunia bisnis konvensional untuk merambah ke dunia bisnis digital. Jangkauan pasar yang semakin luas menjadi salah satu alasan maraknya hal ini. Selain itu, model bisnis ini juga bisa menekan besarnya modal awal yang harus diinvestasikan pelaku usaha untuk menjalankan bisnis. Berbagai macam sektor bisnis yang perkembangannya signifikan dengan memanfaatkan teknologi digital misalnya layanan *ride sharing* atau transportasi online, *e-commerce*, jasa ekspedisi online, dan bisnis *catering online*.

Sebagai salah satu rintisan bisnis digital yang bergerak di bidang *catering* makanan, Home Catering Malang tidak hanya dituntut untuk menyajikan makanan dengan cita rasa terbaik, namun juga harus bisa mengantarkan makanan tepat waktu kepada pelanggan. Ini menjadi standar wajib karena akan mempengaruhi kepuasan dan kepercayaan pelanggan terhadap produk dan layanan yang diberikan.

Kepuasan pelanggan menjadi salah satu faktor kunci dalam dunia bisnis, dan salah satu yang menjadi faktor penentu kepuasan pelanggan dalam bisnis *catering* adalah ketepatan waktu pengiriman. Bisa dikatakan, alasan kenapa pelanggan lebih memilih berlangganan atau memesan makanan secara online salah satunya karena mereka tidak ingin disibukkan dengan meluangkan banyak waktu untuk memasak atau hanya sekedar mencari makanan yang cocok di luar.

Mereka ingin yang serba praktis dan menghemat tenaga, apalagi di era digital seperti sekarang yang hampir semua bisa didapatkan hanya dalam satu gengaman. Ya, mereka bisa mencari apapun di dalam ponsel pintar mereka dengan memanfaatkan koneksi internet. Oleh karenanya, ketepatan waktu pengiriman menjadi sangat berpengaruh terhadap loyalitas dan keberlanjutan pelanggan di bisnis *catering*.

Pengiriman yang efisien juga akan berdampak positif terhadap perkembangan bisnis. Pesanan yang lebih cepat sampai ke pelanggan tentunya akan memberi ruang lebih bagi pemilik bisnis untuk mengembangkan usahanya. Misalnya, dengan pengiriman yang cepat pemilik bisnis bisa memanfaatkan waktu yang tersisa untuk melakukan *quality control* terhadap produknya, melakukan promosi dan membuat perencanaan berikutnya, serta yang paling signifikan adalah manajemen pelanggan dan sumber daya atau tenaga kerja. Dengan waktu pengiriman yang selalu terkontrol dengan baik, pemilik bisnis akan mudah membuat perencanaan dan keputusan apabila ada penambahan jumlah pelanggan. Misalnya dengan jumlah kurir yang tersedia apakah cukup untuk menangani jumlah pelanggan yang ada agar makanan tetap terkirim tepat waktu? Jika belum pemilik bisnis bisa mengatasinya dengan menambah kurir dan menata ulang proses pengiriman. Pemilik bisnis harus benar-benar teliti dalam memperkirakan lama waktu pengiriman.

Manajemen waktu menjadi kunci pada sebagian besar bisnis ataupun pekerjaan manusia. Setiap orang harus seefektif mungkin dalam mengatur waktu yang dimiliki. Karena dengan memanfaatkan waktu dengan baik, maka akan membuka peluang untuk merencanakan dan menyelesaikan masalah atau pekerjaan

selanjutnya dengan lebih cepat pula. Sehingga memberi peluang kepada diri untuk selalu berkembang, baik dalam hal pekerjaan maupun secara pribadi sebagai manusia ciptaan Tuhan. Sebagaimana yang terkandung dalam Al-Quran surat Asy-Syarah ayat 6-8:

إِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٦﴾ فَإِذَا فَرَغْتَ فَانصَبْ ﴿٧﴾ وَإِلَىٰ رَبِّكَ فَارْغَبْ ﴿٨﴾

Artinya : *“Maka sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras (untuk urusan yang lain). Dan hanya kepada Tuhanmu lah engkau berharap.”*

Potongan surat tersebut mengajarkan betapa setiap orang harus bergegas dengan urusannya. Karena masih banyak hal lain yang menanti untuk segera dihadapi dan diselesaikan. Demikian pula dengan usaha *catering* ini, semakin cepat menyelesaikan proses pengiriman, maka pelaku usaha akan mendapat ruang lebih banyak untuk membuat perencanaan dan mengerjakan hal lain untuk lebih memajukan bisnisnya.

Sejauh ini proses pengiriman di Home Catering Malang belum bisa dikatakan optimal. Hal ini tercermin dari data komplain pelanggan yang mayoritas diisi oleh keluhan keterlambatan pengiriman. Dari 71 data komplain yang dihimpun dari bulan Januari sampai dengan Maret 2018, 54 diantaranya (atau sekitar 76 persen) adalah tentang pengiriman yang tidak tepat waktu. Ini tentu sangat berpengaruh terhadap kepuasan pelanggan. Dan bukan tidak mungkin, dalam jangka waktu yang panjang hal ini akan berpengaruh terhadap keberlangsungan bisnis Home Catering Malang.

Untuk itu, adanya sistem yang mampu menganalisa lama waktu pengiriman *catering* diharapkan dapat memberikan informasi akurat kepada pemilik bisnis

dalam memprediksi waktu yang dibutuhkan untuk menyelesaikan proses pengiriman kepada seluruh pelanggan. Sistem ini akan memanfaatkan data statistik pengiriman harian yang telah disediakan oleh Home Catering Malang sebagai data *training*. Beberapa atribut yang mempengaruhi lama pengiriman diantaranya adalah daerah atau rute pengiriman, jumlah pelanggan dalam satu rute, dan total jarak tempuh.

Agar sistem yang dibangun nantinya bisa memberikan informasi yang memiliki nilai akurasi tinggi, penulis menerapkan konsep data mining dengan menggunakan metode regresi linear untuk mengolah data. Dimana hasilnya akan berupa formula atau persamaan linear yang dapat digunakan untuk memproses data baru (data *testing*) untuk memprediksi lama waktu pengiriman.

1.2 Masalah Penelitian

Seberapa besar akurasi metode regresi linear dalam memprediksi lama waktu pengiriman *catering* kepada pelanggan?

1.3 Tujuan Penelitian

Mengukur akurasi metode regresi linear untuk memprediksi lama waktu pengiriman *catering* kepada pelanggan.

1.4 Batasan Masalah

1. Variabel bebas sebagai faktor yang mempengaruhi lama waktu pengiriman adalah jumlah pelanggan dalam satu rute dan total jarak tempuh (meter).
2. *Training* data dibagi menjadi perwilayah atau perdaerah pengiriman, sehingga nantinya setiap daerah akan memiliki persamaan regresi sendiri

sesuai dengan data jumlah pelanggan dan total jarak pada masing-masing wilayah. Daerah pengiriman sendiri terdiri dari area Soekarno-Hatta, Dinoyo, Tlogomas, Sumbersari, dan BCT.

3. Hasil akhir dari penelitian ini berupa prediksi lama waktu yang dibutuhkan untuk menyelesaikan pengiriman berdasarkan parameter-parameter yang diinputkan ke dalam persamaan linear yang merupakan formula matematis yang terbentuk dari hasil analisis regresi linear terhadap data *training*.

1.5 Manfaat Penelitian

Dengan adanya sistem yang bisa memprediksi lama waktu pengiriman *catering* ini, diharapkan dapat memberikan gambaran ilmiah tentang proses prediksi lama waktu pengiriman dengan metode regresi linear terhadap data yang diolah.

BAB 2

KAJIAN PUSTAKA

2.1 Landasan Teori

2.1.1 Prediksi

Prediksi atau ramalan pada dasarnya merupakan dugaan atau prediksi mengenai terjadinya suatu kejadian atau peristiwa di waktu yang akan datang. Prediksi bisa bersifat kualitatif (tidak berbentuk angka) maupun kuantitatif (berbentuk angka). Prediksi kualitatif sulit dilakukan untuk memperoleh hasil yang baik karena variabelnya sangat relatif sifatnya. Prediksi kuantitatif dibagi dua yaitu: prediksi tunggal (*point prediction*) dan prediksi selang (*interval prediction*). Prediksi tunggal terdiri dari satu nilai, sedangkan prediksi selang terdiri dari beberapa nilai, berupa suatu selang (interval) yang dibatasi oleh nilai batas bawah (prediksi batas bawah) dan batas atas (prediksi tinggi) (Douglas C Montgomery & Team, 2013).

Prediksi berfungsi untuk membuat suatu rencana kebutuhan (*demand*) yang harus dibuat yang dinyatakan dalam kuantitas (jumlah) sebagai fungsi dari waktu. Prediksi dilakukan dalam jangka panjang (*long term*). Prediksi yang berkaitan dengan pernyataan (1) *what will be demanded*, (2) *how many*, dan (3) *when it should be supplied* ? Prediksi sangat diperlukan dengan melakukan perbandingan antara kebutuhan yang diramalkan dengan yang sebenarnya.

a) Prediksi Kualitatif

Yaitu prediksi yang didasarkan atas kualitatif pada masa lalu. Hasil prediksi yang dibuat sangat tergantung pada orang yang menyusunnya. Hal ini penting karena hasil prediksi tersebut ditentukan berdasarkan pemikiran yang bersifat intuisi, judgement atau pendapat, dan pengetahuan serta pengalaman dari penyusunnya. Biasanya prediksi secara kualitatif ini didasarkan atas hasil penyelidikan.

Beberapa metode prediksi yang digolongkan sebagai model kualitatif adalah sebagai berikut :

- 1) Metode *Delphi*, Sekelompok pakar mengisi kuesioner, merupakan teknik prediksi berdasarkan pada proses konvergensi dari opini beberapa orang atau ahli secara interaktif tanpa menyebutkan identitasnya. Moderator menyimpulkan hasilnya dan memformulasikan menjadi suatu kuesioner baru yang diisi kembali oleh kelompok tersebut, demikian seterusnya. Hal ini merupakan proses pembelajaran (*learning process*) dari kelompok tanpa adanya tekanan atau intimidasi individu.
- 2) Dugaan manajemen atau Panel Consensus, dimana prediksi semata-mata berdasarkan pertimbangan manajemen, umumnya oleh manajemen senior. Metode ini akan cocok dalam situasi yang sangat sensitif terhadap intuisi dari suatu atau sekelompok kecil orang yang karena pengalamannya mampu memberikan opini yang kritis dan relevan. Teknik akan dipergunakan dalam situasi dimana tidak ada situasi dimana tidak ada alternatif lain dari model prediksi yang

dapat diterapkan. Bagaimanapun metode ini mempunyai banyak keterbatasan, sehingga perlu dikombinasikan dengan metode prediksi yang lain.

b) **Prediksi Kuantitatif**

Yaitu prediksi yang didasarkan atas data kuantitatif pada masa lalu. Hasil prediksi yang dibuat sangat tergantung pada metode yang dipergunakan dalam prediksi tersebut. Dengan metode yang berbeda akan diperoleh hasil prediksi yang berbeda, adapun yang perlu diperhatikan dari penggunaan metode tersebut, adalah baik tidaknya metode yang dipergunakan, sangat ditentukan oleh perbedaan atau penyimpangan antara hasil ramalan dengan kenyataan yang terjadi. Metode yang baik adalah metode yang memberikan nilai-nilai perbedaan atau penyimpangan yang mungkin. Prediksi kuantitatif hanya dapat digunakan apabila terdapat tiga kondisi sebagai berikut:

- 1) Tersedia informasi tentang masa lalu (data historis).
- 2) Informasi tersebut dapat dikuantitatifkan dalam bentuk numerik.
- 3) Dapat diasumsikan bahwa beberapa aspek pola masa lalu akan terus berlanjut di masa mendatang.

Variabel keputusan menandakan kehadiran sesuatu yang dapat diukur. Pengukuran dapat dilakukan bila sebuah model menghasilkan atau dibangun atas Nilai numeris dari setiap variabel keputusan.

Nilai numeris ini mengimplikasikan terdapatnya keputusan yang khusus. Keputusan adalah angka, seluruh model yang dibangun terdiri dari

angka numeris dan huruf. Keduanya sama-sama digunakan, dimana angka lebih dominan penggunaannya.

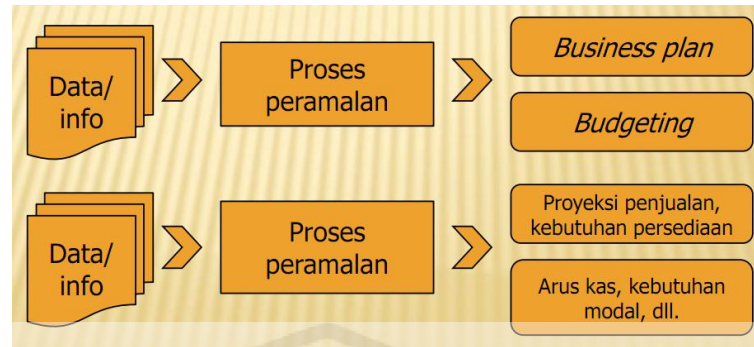
Beberapa metode prediksi yang digolongkan sebagai model kuantitatif adalah sebagai berikut :

1) Regresi dan Korelasi

Metode regresi dan korelasi pada penetapan suatu persamaan estimasi menggunakan teknik *least squares*. Hubungan yang ada pertama-tama dianalisis secara statistik. Ketepatan prediksi dengan menggunakan metode ini sangat baik untuk prediksi jangka pendek, sedangkan untuk prediksi jangka panjang ternyata ketepatannya kurang begitu baik. Metode ini banyak digunakan untuk prediksi penjualan, perencanaan keuntungan, prediksi permintaan dan permalan keadaan ekonomi.

2) Metode Ekonometrik

Metode ini didasarkan atas prediksi sistem persamaan regresi yang diestimasi secara simultan. Baik untuk prediksi jangka pendek maupun prediksi jangka panjang, ketepatan prediksi dengan metode ini sangat baik. Metode prediksi ini selalu dipergunakan untuk prediksi penjualan menurut kelas produk, atau prediksi keadaan ekonomi masyarakat, seperti permintaan, harga dan penawaran. Data yang dibutuhkan untuk penggunaan metode prediksi ini adalah data kuartalan beberapa tahun.



Gambar 2.1 Gambaran proses prediksi

2.1.2 *Data Mining*

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (McLeod, 1995). *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Kusrini, 2009). *Data mining* sering juga disebut sebagai *Knowledge Discovery In Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, histori untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Susanto, 2010).

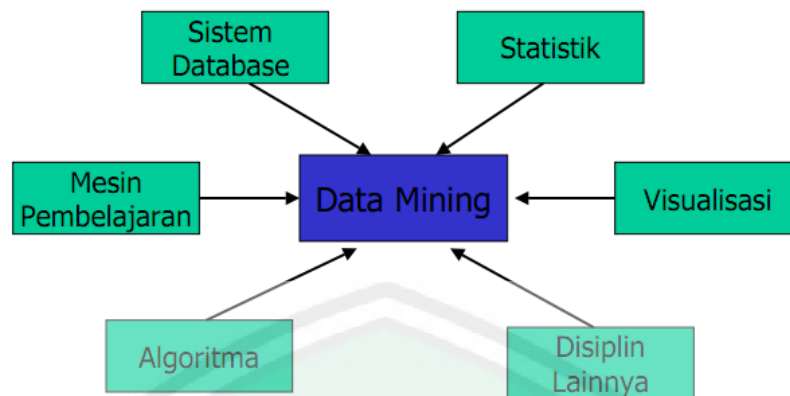
Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, *data warehouse*, *statistic*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti *neural network*, pengenalan pola, spasial data analisis, *image database*, *signal processing* (Turban, 2005).

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis (Turban, 2005). Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi.

Karakteristik *data mining* sebagai berikut :

- a) *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b) *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- c) *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi (McLeod, 1995).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. *Data mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan (*Artificial Intelligent*), *Machine Learning*, Statistik dan *Database*. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *Clustering*, *Classification*, *Association Rules Mining*, *Neural Network*, *Generic Algorithm* dan lain-lain (McLeod, 1995).



Gambar 2.2 Disiplin ilmu dalam *data mining*

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*.

Tahap-tahap data mining ada 7 yaitu :

1) Pembersihan Data (*Data Cleansing*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah kompleksitasnya.

2) Integrasi Data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

3) Seleksi Data (*Data Selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4) Transformasi Data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5) Proses *Mining*

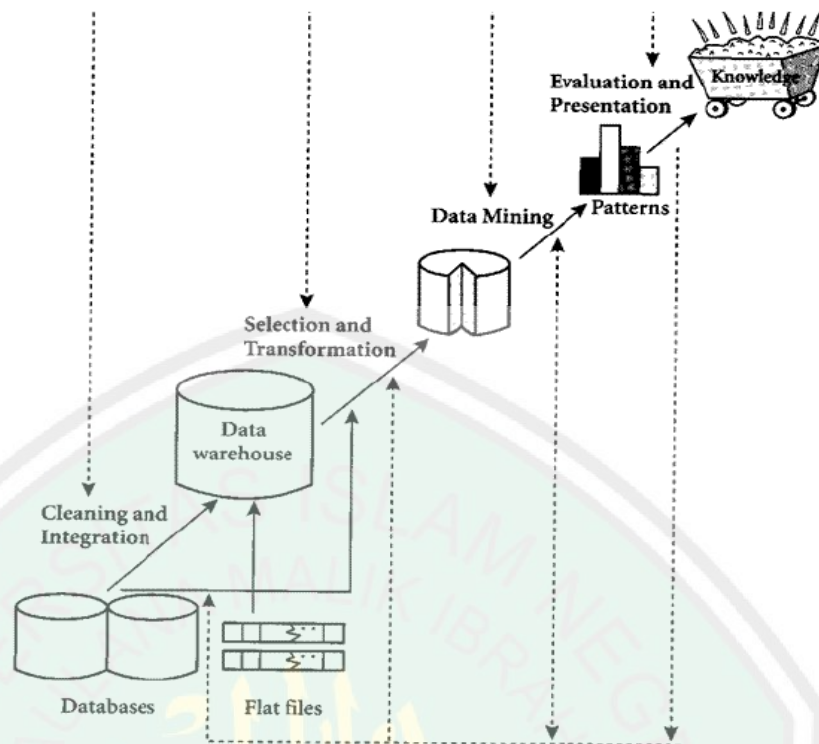
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6) Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7) Presentasi Pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan aksi dari hasil analisis yang didapat. Adakalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining* (Turban, 2005).



Gambar 2.3 Proses *data mining*

2.1.3 Metode Regresi Linear

Regresi linear adalah alat statistik yang dipergunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Variabel yang mempengaruhi sering disebut variabel bebas, variabel independen atau variabel penjelas. Variabel yang dipengaruhi sering disebut dengan variabel terikat atau variabel dependen. Regresi linear hanya dapat digunakan pada skala interval dan rasio.

Regresi sebagai alat ukur yang juga digunakan untuk mengukur ada atau tidaknya korelasi antar variabel. Analisis regresi adalah sebuah teknik statistik untuk membuat model dan menyelidiki hubungan diantara dua atau lebih variabel yang dimaksud diatas. Regresi mengukur seberapa besar suatu

variabel mempengaruhi variabel yang lain, sehingga dapat digunakan untuk melakukan prediksi nilai suatu variabel berdasarkan variabel lain.

Secara umum regresi linear terdiri dari dua, yaitu regresi linear sederhana yaitu dengan satu buah variabel bebas dan satu buah variabel terikat dan regresi linear berganda dengan beberapa variabel bebas dan satu buah variabel terikat. Analisis regresi linear merupakan metode statistik yang paling jamak dipergunakan dalam penelitian-penelitian sosial, terutama penelitian ekonomi (Lungan, 2006).

2.1.4 Regresi Linear Sederhana

Analisis regresi linear sederhana dipergunakan untuk mengetahui pengaruh antara satu buah variabel bebas terhadap satu buah variabel terikat.

Persamaan umumnya adalah:

$$Y = b_0 + b X$$

Dengan Y adalah variabel terikat dan X adalah variabel bebas. Koefisien b_0 adalah konstanta (*intercept*) yang merupakan titik potong antara garis regresi dengan sumbu Y pada koordinat kartesius.



Gambar 2.4 Garis regresi Y karena pengaruh X, persamaan regresinya $Y=2,0 + 0,5 X$

2.1.5 Regresi Linear Berganda

Regresi linear berganda merupakan perluasan dari regresi linear sederhana. Jika regresi linear sederhana mempersoalkan tentang hubungan variabel terikat atau variabel kriteria (*respons*) dengan satu variabel bebas atau prediktor (deterministik), maka pada regresi linear berganda mempersoalkan hubungan linear antara satu variabel terikat dengan beberapa variabel bebas.

Variabel terikat dapat berupa ukuran atau kriteria keberhasilan, sedangkan variabel bebas dapat berupa factor-faktor penentu keberhasilan tersebut. Misalkan banyaknya buah ditentukan oleh dosis pemupukan, iklim, tanah, dan air (faktor-faktor penentu banyaknya buah).

Persamaan regresi untuk dua prediktor adalah :

$$Y = a + b_1X_1 + b_2X_2$$

Persamaan regresi untuk tiga prediktor adalah :

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

Persamaan regresi untuk n prediktor adalah :

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

Untuk bisa membuat ramalan melalui regresi, maka data setiap variabel harus tersedia. Selanjutnya berdasarkan data itu peneliti harus dapat menemukan persamaan regresi melalui perhitungan.

2.1.6 Asumsi Klasik dalam Regresi Linear

Asumsi Klasik merupakan salah satu pengujian prasyarat pada regresi linear berganda. Suatu model regresi yang valid harus memenuhi kriteria *BLUE* (*Best, Linear, Unbiased, and Estimated*). Untuk dapat mengetahui apakah model regresi yang kita gunakan dalam penelitian telah memenuhi kriteria *BLUE*, maka dilakukan uji prasyarat regresi linear berganda, yaitu uji Asumsi Klasik (Kuncoro, 2013).

Asumsi Klasik pada umumnya terdiri dari berbagai pengujian. Namun, umumnya yang sering digunakan Uji Multikolinearitas, Autokorelasi, Normalitas, dan Heteroskedastisitas (Robert Kurniawan, 2016).

2.1.6.1 Multikolinearitas

Hubungan linear antarvariabel bebas disebut dengan multikolinearitas. Hubungan tersebut tercipta karena adanya korelasi antarvariabel bebas, di mana setiap ada perubahan pada suatu variabel

bebas, akan mengakibatkan variabel bebas lainnya berubah. Oleh karena itu, dalam membuat regresi berganda, variabel bebas yang baik adalah variabel bebas yang mempunyai hubungan dengan variabel terikat, tetapi tidak mempunyai hubungan dengan variabel bebas lainnya.

Dalam praktiknya, umumnya multikolinieritas tidak dapat dihindari. Dalam artian, sulit menemukan dua variabel bebas yang secara matematis tidak berkorelasi (korelasi = 0) sekalipun secara substansi tidak berkorelasi. Akan tetapi, ada multikolinieritas yang signifikan (harus mendapat perhatian khusus) dan tidak signifikan (mendekati nol). Kita juga akan sulit menemukan kolinieritas yang sempurna (*perfect collinierity*). Bila ditemukan kolinieritas yang sempurna, maka salah satu dampak yang ditimbulkannya adalah tidak dapat dihitungnya koefisien regresi.

Secara umum, bila terjadi kolinieritas sempurna atau terjadi korelasi yang tinggi antarvariabel bebas akan menyebabkan nilai determinan dari matriks $x'x$ akan mendekati 0 (akan sama dengan 0 jika korelasi sempurna), maka koefisien regresi tidak dapat diperoleh dikarenakan $(x'x)^{-1}$ tidak dapat dicari. Satu hal yang perlu ditekankan kembali di sini bahwa multikolinieritas merupakan hubungan linear. Jika variabel bebas mempunyai hubungan, tetapi tidak linear maka kondisi tersebut tidak dapat dikategorikan sebagai multikolinieritas, misalnya $x_1 = x_2^2$ atau $x_1 = \log x_2$, maka dikatakan bahwa x_1 dan x_2 tidak kolinier.

a) Dampak Multikolinieritas

Secara implisit sebenarnya pemaparan di atas telah menunjukkan beberapa dampak yang ditimbulkan oleh multikolinieritas, seperti: tidak

dapatnya dilakukan interpretasi, atau tidak dapatnya koefisien regresi untuk diestimasi. Dalam praktiknya, kolinieritas sempurna hampir tidak ditemui, sehingga sekalipun variabel bebas berkorelasi, koefisien regresi tetap dapat diestimasi. Akan tetapi, dengan terdapatnya multikolinearitas dalam persamaan regresi, maka membawa berbagai konsekuensi terhadap model itu sendiri.

Ada beberapa dampak yang ditimbulkan oleh kolinieritas tersebut, antara lain:

1. Varian koefisien regresi menjadi besar.
2. Koefisien Partial Regresi tidak terukur secara presisi. Oleh karena itu nilai standar errornya besar.
3. Perubahan kecil pada data dari sampel ke sampel akan menyebabkan perubahan drastis pada nilai koefisien regresi partial.
4. Perubahan pada satu variabel dapat menyebabkan perubahan besar pada nilai koefisien regresi parsial variabel lainnya.
5. Nilai *Confidence Interval* sangat lebar, sehingga akan menjadi sangat sulit untuk menolak hipotesis nol pada sebuah penelitian jika dalam penelitian tersebut terdapat multikolinearitas.

b) Teknik Mendeteksi Multikolinearitas

Ada banyak uji formal untuk mendeteksi keberadaan multikolinearitas. Salah satu metode yang paling sederhana adalah dengan menggunakan rumus korelasi *Pearson* untuk mengetahui seberapa besar nilai R antarvariabel bebas namun kita seringkali dihadapkan pada keraguan seperti berapa besarkah nilai R yang dianggap besar. Karena ada sumber

yang mengatakan 70%, 80% atau 90%. Batasan tersebut jadi bersifat substansial dan subjektif.

Uji formal lain yang populer untuk mendeteksi keberadaan multikolinearitas adalah *VIF* dan *Tolerance*.

1. *VIF (Variance Inflation Factor)*

Apabila nilai *VIF* besar menunjukkan adanya kolinearitas antara variabel x_1 dan x_2 . Dengan kata lain, apabila *VIF* kecil, maka dapat diduga tidak ada multikolinearitas. Berdasarkan besarnya koefisien determinasi, antar variabel bebas tidak berkorelasi apabila nilai koefisien determinasinya sama dengan 0 ($R_j^2 = 0$). Dengan R_j^2 bernilai 0 maka nilai *VIF* = 1. Sebaliknya, jika $R_j^2 \neq 0$ maka nilai *VIF* > 1. Dapat disimpulkan bahwa, kolinearitas tidak ada apabila nilai *VIF* mendekati angka 1. Salah satu sumber baku mengatakan bahwa kolinearitas dianggap ada jika nilai *VIF* > 5, artinya nilai $R_j^2 = 0,8$. Namun ada standar lain berpendapat bahwa ada multikolinearitas ketika nilai *VIF* > 10 atau nilai $R_j^2 = 0,9$.

2. *Tolerance (TOL)*

Tolerance merupakan kebalikan dari *VIF*. Jika $R_j^2 = 0$ atau antarvariabel bebas tidak ada korelasi, maka *TOL* = 1, sebaliknya jika $R_j^2 = 1$ atau mempunyai korelasi sempurna maka nilai *TOL* = 0. Oleh karena itu, berdasarkan aturan ini variabel bebas dinyatakan tidak multikolinearitas jika *TOL* mendekati 1.

c) Mengatasi Multikoliniearitas

Beberapa alternatif cara untuk mengatasi masalah multikolinearitas, sebagai berikut:

1. Mengganti atau mengeluarkan variabel yang mempunyai korelasi yang tinggi. Namun hal ini harus dilakukan secara hati-hati, karena dapat menyebabkan variabel yang dikeluarkan merupakan variabel yang secara teoretis penting. Kondisi ini dikenal dengan sebutan *specification bias*.
2. Menambah jumlah observasi. Dengan menambah jumlah data diharapkan kolinearitas dapat berkurang karena terkadang sampel lain tidak memiliki kasus multikolinearitas yang serius. Sampel yang lebih besar selalu akan meningkatkan tingkat akurasi atau presisi estimasi B , namun jumlah sampel yang besar tidak bisa menghilangkan kesulitan yang muncul pada saat proses interpretasi koefisien regresi ketika terjadi multikolinearitas yang kuat.
3. Mentransformasikan data ke dalam bentuk lain, misalnya logaritma natural, akar kuadrat atau bentuk first difference dan ratio transformation method. Namun harus diperhatikan bahwa dengan melakukan transformasi bukan berarti tidak menyebabkan masalah lain, seperti tidak terpenuhinya asumsi model regresi linear klasik, sampai terjadinya heteroskedastisitas.
4. Menggunakan metode-metode regresi yang lebih advance seperti Ridge Regression.

2.1.6.2 Heteroskedastisitas

Dalam analisis regresi, biasanya akan ada metode yang digunakan untuk melakukan estimasi parameter. Salah satu metode yang paling sering digunakan adalah Metode Kuadrat Terkecil (*Ordinary Least Square*). Prinsip dari metode tersebut yaitu dengan meminimumkan jumlah kuadrat galat, atau biasa disebut residual. Tetapi, ada beberapa asumsi di dalam analisis regresi yang harus dipenuhi untuk melakukan estimasi dengan metode OLS tersebut. Beberapa asumsi tersebut, di antaranya:

1. Data harus mengikuti sebaran normal.
2. Tidak ada multikolinearitas.
3. Tidak ada autokorelasi pada data.
4. Data bersifat homoskedastik.

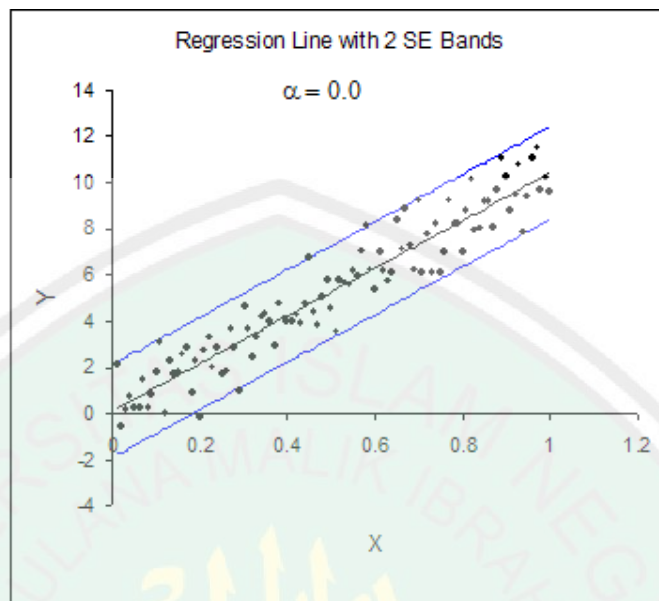
Jika asumsi di atas terpenuhi semua, maka hasil estimasi menggunakan metode kuadrat terkecil akan memenuhi sifat *Best, Linear, Unbiased, Estimator (BLUE)*. Akan tetapi, jika ada satu saja, atau bahkan lebih dari satu, asumsi yang tidak terpenuhi, maka hasil estimasi yang diperoleh tidak dapat memenuhi sifat *BLUE* tersebut.

Di antara asumsi-asumsi tersebut, ada asumsi penting yang harus dipenuhi, yaitu asumsi homoskedastisitas. Homoskedastisitas memiliki kebalikan, yaitu heteroskedastisitas. Heteroskedastisitas adalah adanya ketidaksamaan varian dari residual untuk semua pengamatan pada model regresi. Heteroskedastisitas secara simbolis dapat ditulis seperti berikut:

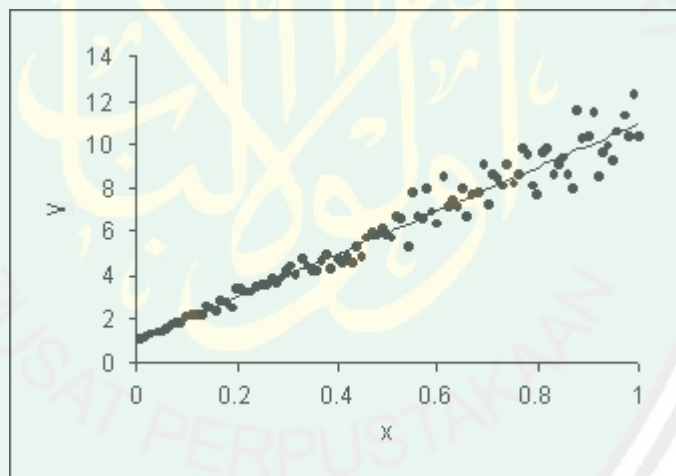
$$\text{Var}(e|x_1, x_2, \dots, x_k) = \delta^2$$

Jika dilihat pada grafik, heteroskedastisitas dapat dilihat jika persebaran data menjadi tidak tetap seiring berjalannya waktu. Adapun persebaran data

jika terjadi homoskedastisitas adalah tetap. Mungkin lebih jelasnya akan diberikan gambar contoh.



Gambar 2.5 Contoh Homoskedastisitas



Gambar 2.6 Contoh Heteroskedastisitas

Terlihat bahwa pada gambar pertama persebaran data tetap seiring waktu. Adapun pada gambar yang kedua, persebaran data berubah seiring waktu. Tentunya keadaan heteroskedastisitas tidak diharapkan di dalam model, karena kita akan kesulitan untuk mengestimasi model yang tepat akibat varian data yang tidak konsisten.

a) Dampak Heteroskedastisitas

Sebenarnya penyimpangan asumsi homoskedastisitas atau yang kita sebut sebelumnya dengan heteroskedastisitas terhadap operasi OLS kemungkinan tidak akan merusak sifat *unbiased* dan konsistensinya, namun berdampak besar dalam hal lain. Dampak dari penyimpangan asumsi homoskedastisitas, yaitu:

1. Heteroskedastisitas Merusak Efisiensi Estimator OLS

Rusaknya sifat efisiensi estimator OLS tersebut menyebabkan hasil pengujian hipotesisnya menjadi meragukan.

2. Heteroskedastisitas Merusak Varians

Heteroskedastisitas menyebabkan varians yang dihasilkan cenderung membesar sehingga tidak lagi merupakan varian yang terkecil. Kecenderungan semakin membesarnya varian tersebut akan mengakibatkan uji hipotesis yang dilakukan juga tidak akan memberikan hasil yang baik (tidak valid). Selain itu, varians yang membesar juga mengakibatkan selang kepercayaan yang melebar. Tentu hal ini tidak baik dalam pemilihan model regresi.

b) Teknik Mendeteksi Heteroskedastisitas

Biasanya heteroskedastisitas terjadi pada data cross section, yaitu data yang diambil pada satu waktu, yang mewakili berbagai ukuran (kecil, sedang, dan besar). Oleh karena itu, cara paling mudah mendeteksi heteroskedastisitas adalah dengan melihat grafik atau gambar. Adapun cara lain untuk mendeteksi ada tidaknya heteroskedastisitas yaitu dengan melakukan beberapa pengujian, diantaranya adalah:

1. Uji *White*
2. Uji Korelasi *Spearman*
3. Uji *Park*

Adapun beberapa metode yang digunakan untuk mengatasi heteroskedastisitas diantaranya :

1. Transformasi Logaritma Normal
2. Pemilihan model terbaik
3. Menggunakan *Robust Standar Error*
4. Menggunakan *Weighted Least Square*

2.1.6.3 Autokorelasi

Secara garis besar, uji autokorelasi digunakan untuk mendeteksi apakah terjadi korelasi antara residu pada periode saat ini (t) dengan residu pada periode satu periode sebelumnya ($t-1$). Untuk memenuhi kriteria *BLUE*, model regresi harus terbebas dari gejala autokorelasi. Khususnya masalah autokorelasi cenderung terjadi pada penelitian dengan menggunakan data *time series*, sementara itu sangat jarang terjadi pada penelitian dengan menggunakan data *cross section*. Untuk mendeteksi gejala autokorelasi dapat menggunakan uji statistik yaitu uji *Durbin-Watson* dengan kriteria pengambilan keputusan sebagai berikut (Algifari, 2000):

1. Apabila $dw < dl$, maka terjadi autokorelasi negatif
2. Apabila $dl < dw < du$, maka tidak dapat disimpulkan
3. Apabila $du < dw < 4-du$, maka tidak terjadi autokorelasi negatif dan positif
4. Apabila $4-du > dw > 4-dl$, maka tidak dapat disimpulkan

5. Apabila $4-dl < dw$, maka terjadi autokorelasi positif

2.1.6.4 Normalitas

Uji Normalitas data merupakan pengujian asumsi klasik paling utama yang harus dilakukan oleh peneliti. Dalam melakukan penelitian, data harus mendekati distribusi normal. Tujuan uji normalitas adalah untuk menguji apakah dalam model regresi, variabel dependen dan variabel independen mempunyai distribusi normal atau tidak. Kenormalan suatu data merupakan syarat wajib suatu yang harus terpenuhi dalam model regresi linear (R Adisetiawan, 2011). Salah satu cara untuk mengetahui apakah data penelitian berdistribusi normal atau tidak dapat melihat *normal probability plots* (Ghozali, 2005). Pengujian ini dilakukan dengan cara melihat titik-titik penyebaran data terhadap garis diagonal pada grafik. Kriteria pengambilan keputusan analisis *normal probability plots* adalah sebagai berikut:

1. Apabila data (yang dapat dilihat dari titik-titik pada grafik) menyebar dan mengikuti garis diagonal, maka dapat disimpulkan bahwa data mengikuti pola distribusi normal.
2. Apabila data (yang dapat dilihat dari titik-titik pada grafik) menyebar dan cenderung menjauh dari garis diagonal serta tidak mengikuti garis diagonal, maka dapat disimpulkan data tidak menunjukkan pola distribusi normal.

2.1.7 Akurasi Prediksi

Dalam semua situasi prediksi itu mengandung derajat ketidakpastian. Kita mengenal fakta ini dengan memasukkan unsur error

atau kesalahan dalam perumusan sebuah prediksi deret waktu. Sumber penyimpangan dalam prediksi bukan hanya disebabkan oleh unsur error tetapi ketidakmampuan suatu model prediksi mengenali unsur yang lain dalam deret data yang mempengaruhi besarnya penyimpangan dalam prediksi.

Jadi besarnya penyimpangan hasil prediksi dapat disebabkan oleh besarnya faktor yang tidak diduga (*outliers*) dimana tidak ada metode prediksi yang mampu menghasilkan prediksi yang akurat atau dapat juga disebabkan metode prediksi yang digunakan tidak dapat memprediksi dengan tepat komponen tren, komponen musiman atau komponen siklus yang mungkin terdapat dalam deret data dan itu berarti metode yang digunakan tidak tepat (Bowerman, 1993).

Dalam permodelan deret berkala sebagian data diketahui dapat dipergunakan untuk memprediksi sisa data berikutnya sehingga dapat dilakukan perhitungan ketepatan prediksi secara baik. Ketepatan prediksi dimasa yang akan datang adalah sangat penting. Ukuran akurasi hasil prediksi yang merupakan ukuran kesalahan prediksi merupakan ukuran tentang tingkat perbedaan antara hasil prediksi dengan permintaan yang sebenarnya terjadi. Ukuran yang bisa digunakan, yaitu :

1. Rata-Rata Deviasi Mutlak

Mean Absolute Deviation (MAD) merupakan rata-rata kesalahan mutlak selama perioda tertentu tanpa memperhatikan apakah hasil permalalan lebih besar atau lebih kecil dibandingkan kenyataannya. Secara matematis, *Mean Absolute Deviation* dirumuskan sebagai berikut :

$$MAD = \frac{\sum |A_t - F_t|}{n}$$

Keterangan :

MAD = *Mean Absolute Deviation*

F_t = Prediksi untuk periode t

A_t = Nilai Aktual

n = Jumlah data

2. Rata-Rata Persentase Kesalahan Absolut

Mean Absolute Percentage Error (MAPE) merupakan ukuran kesalahan relatif. *Mean Absolute Percentage Error* biasanya lebih berarti dibandingkan *Mean Absolute Deviation* karena *Mean Absolute Percentage Error* menyatakan persentase kesalahan hasil peramalan terhadap permintaan aktual selama periode tertentu yang akan memberikan informasi persentase kesalahan terlalu tinggi atau terlalu rendah. Secara matematis, *Mean Absolute Percentage Error* dinyatakan sebagai berikut:

$$MAPE = \frac{100 \sum_{i=1}^n \left[\frac{Y_i - \hat{Y}_i}{Y_i} \right]}{n}$$

Keterangan :

$MAPE$ = *Mean Absolute Percentage Error*

Y = Nilai aktual

n = Jumlah data

2.2 Profil Home Catering Malang

Home Catering Malang adalah *home industry* yang bergerak di bidang makanan. Sesuai namanya, wirausaha rumahan ini fokus pada *catering* yang

mengandalkan pengiriman untuk mengantar makanan kepada pelanggannya. Berdiri sejak akhir tahun 2017, setiap harinya Home Catering berkewajiban untuk mengirim rata-rata 300 kotak *catering* kepada pelanggan yang tersebar di beberapa wilayah di Kota Malang.



Gambar 2.7 Logo Home Catering Malang

Dalam perkembangannya, Home Catering Malang mulai melirik ke ranah digital. Wirausaha kuliner rumahan ini mulai merintis peluang untuk menjadi sebuah *start up* yang bergerak di bidang industri makanan. Tak hanya memproduksi, namun juga *delivery*. Semua itu tercermin dari beberapa proses bisnisnya yang menggantungkan pada perangkat teknologi dan internet. Mulai dari *marketing*, promosi dan pengembangan bisnis, pembayaran hingga pengiriman kepada pelanggan. Karena pandangan visioner itu pula, Home Catering Malang melihat data sebagai aset yang bisa dianalisa untuk perkembangan bisnis. Mulai dari data pelanggan dan persebarannya, sumber daya, pengiriman, pembayaran, hingga data komplain pelanggan. Semua dicatat secara terstruktur dan dilakukan pembaharuan secara berkala.

2.3 Penelitian Terkait

Beberapa penelitian sebelumnya yang terkait dengan penelitian ini diantaranya :

- 1) Implementasi Metode Regresi Linear Dalam Pembuatan Aplikasi Simulasi Sistem Penentuan Kelayakan Lokasi Pembangunan Perumahan.

Dilakukan oleh Syaharullah Disa pada tahun 2011. Penelitian ini bertujuan untuk membuat aplikasi simulasi komputer yang berfungsi sebagai sistem pendukung keputusan dalam menentukan kelayakan lokasi perumahan. Metode yang digunakan adalah Regresi Linear. Fungsi regresi linear diimplementasikan dalam program simulasi yang dibuat. Dengan menerapkan metode regresi linear dapat diketahui bahwa faktor-faktor yang paling dominan dalam penentuan kelayakan lokasi pembangunan perumahan adalah sarana dan prasarana 98,83%, legalitas atau kepemilikan tanah 88,37%, lingkungan 87,79%, 84,30%, kelayakan fisik 80,81%, aksesibilitas 76,74%. (Disa, 2011)

- 2) Prediksi Tingkat Produksi Kopi Menggunakan Regresi Linear.

Dilakukan oleh Petrus Katemba dan Rosita Koro Djoh (2017). Penelitian bertujuan untuk menganalisa dan melakukan prediksi produksi kopi di Kabupaten Manggarai. Selanjutnya dilakukan pengujian menggunakan MSE dan MAPE, hasilnya diperoleh nilai MSE 43,112% dan MAPE 20,001% sehingga pengujian menggunakan MAPE jauh lebih baik dalam menghitung akurasi prediksi produksi kopi. Dengan memanfaatkan metode regresi linear, penelitian ini mampu memprediksi tingkat produksi

kopi di tahun tertentu berdasarkan data-data hasil produksi di tahun sebelumnya, yakni dalam rentang tahun 2011-2015 (Koro, 2017).



BAB 3

METODE PENELITIAN

3.1 Sumber Data

Pada penelitian ini terdapat sejumlah data yang digunakan sebagai data latih (*training*). Data ini berupa data pengiriman harian oleh setiap kurir Home Catering dalam rentang waktu satu setengah bulan, yaitu mulai tanggal 15 Maret sampai 30 April 2018. Terdapat beberapa atribut diantaranya tanggal pengiriman, nama kurir, jumlah pelanggan, daerah atau rute pengiriman, jarak tempuh, dan lama waktu pengiriman.

Data yang digunakan dalam penelitian ini berupa data primer yang diperoleh langsung dari Home Catering Malang. Data tersebut berupa data pengiriman harian yang kemudian dilakukan proses *preprocessing* agar bisa dimanfaatkan untuk keperluan analisa dengan teknik *data mining*, sehingga dapat merumuskan lama waktu pengiriman jika terjadi perubahan nilai variabel yang dapat mempengaruhi waktu pengiriman. Total data *training* yang akan digunakan pada penelitian ini yaitu sebanyak 125 *records*.

Karena perbedaan karakteristik tiap daerah yang bisa mempengaruhi lama waktu pengiriman berbeda-beda, maka data *training* akan dipecah menjadi lima kelompok (sesuai dengan jumlah daerah yang menjadi basis pelanggan Home Catering Malang). Jadi setiap daerah nantinya akan memiliki persamaan regresi masing-masing. Tujuannya adalah untuk meminimalisir faktor atau pengaruh di luar persamaan yang dapat menurunkan nilai koefisien determinasi variabel x terhadap nilai Y .

Sehingga diharapkan dapat meningkatkan akurasi data hasil prediksi lama waktu pengiriman (Y) yang diukur berdasarkan nilai koefisien determinasi (R^2). Persentase nilai R^2 menunjukkan kuatnya hubungan variabel x terhadap Y pada persamaan regresi linear.

Tabel 3.1 Data sampel pengiriman harian

Tanggal	Kurir	Σ Pelanggan	Jarak Tempuh (m)	Daerah	Waktu Kirim (menit)
1-Jan-18	Irfan	25	5000	BCT	90
1-Jan-18	Yusron	28	4500	Suhat	95
1-Jan-18	Pak Sen	26	4300	Tlogomas	90
1-Jan-18	Bagus	28	3000	Dinoyo	60

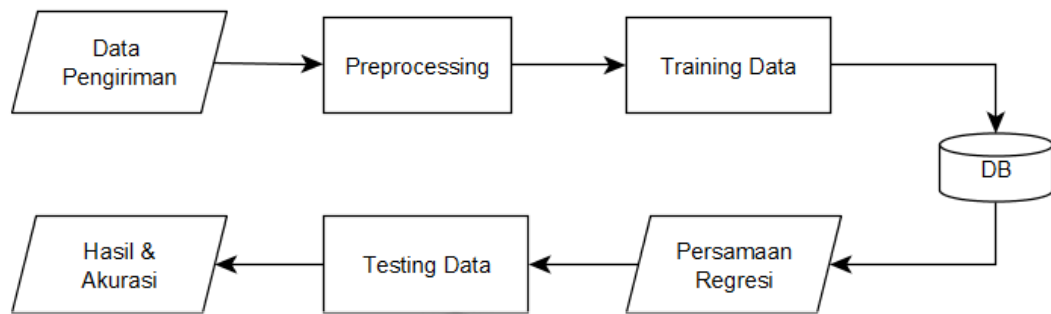
3.2 Objek Penelitian

Lama waktu pengiriman makanan di Home Catering Malang menjadi fokus utama pada penelitian ini. Terdapat beberapa variabel atau atribut data yang mempengaruhi lama waktu pengiriman. Teknik *data mining* dengan metode regresi linear digunakan untuk mengukur akurasi dan memprediksi lama waktu pengiriman.

3.3 Analisis dan Desain Sistem

3.3.1 Desain Sistem

Gambaran umum desain sistem proses penghitungan waktu pengiriman dengan metode regresi linear, dapat dilihat pada Gambar 3.1



Gambar 3.1 Desain sistem

3.3.1.1 Data Pengiriman

Tahap awal dari perancangan sistem ini adalah proses analisis data. Yaitu data pengiriman harian yang terdiri dari beberapa atribut data atau variabel bebas yang berpengaruh terhadap waktu yang dibutuhkan untuk menyelesaikan proses pengiriman. Variabel bebas (X_1 dan X_2) merupakan data input, sedangkan daerah sebagai pengkategorori. Data ini dimasukkan ke dalam persamaan akhir regresi linear hasil dari proses *training* data pengiriman. Sedangkan Y merupakan *output* yang berupa lama waktu pengiriman dalam satuan menit.

3.3.1.2 Praproses Data

Tahap ini memastikan data telah diseleksi dan dibagi menjadi lima kelompok, dikategorikan berdasarkan daerah pengiriman. Sehingga pengelompokan datanya seperti yang digambarkan berikut:

Tabel 3.2 Model data daerah 1 (Soekarno Hatta)

X1	X2	Y1
25	4000	60
25	4100	62
26	4200	64

26	4300	66
28	4400	68

Tabel 3.3 Model data daerah 2 (Dinoyo)

X1	X2	Y2
25	4000	60
25	4200	64
27	4300	68
27	4250	65
24	3950	58

Tabel 3.4 Model data daerah 3 (Tlogomas)

X1	X2	Y3
26	4100	65
26	4220	67
27	4330	69
27	4420	71
29	4520	73

Tabel 3.5 Model data daerah 4 (Sumpersari)

X1	X2	Y4
28	4050	65
28	4200	69
30	4370	73

30	4300	70
27	4000	63

Tabel 3.6 Model data daerah 5 (BCT)

X1	X2	Y5
25	4800	70
25	4900	72
26	5000	74
26	5100	76
28	5200	78

Dari data pada **Tabel 3.2** kemudian didapatkan vektor Y dan matriks X sebagai berikut :

Vektor Y, didapat dari semua data lama waktu pengiriman,

60

62

64

66

68

Matriks X, yang merupakan kumpulan data X1 dan X2

25 4000

25 4100

26	4200
26	4300
28	4400

Selanjutnya menyisipkan vektor baru yaitu μ dengan nilai konstanta 1 ke dalam matriks X sehingga menjadi,

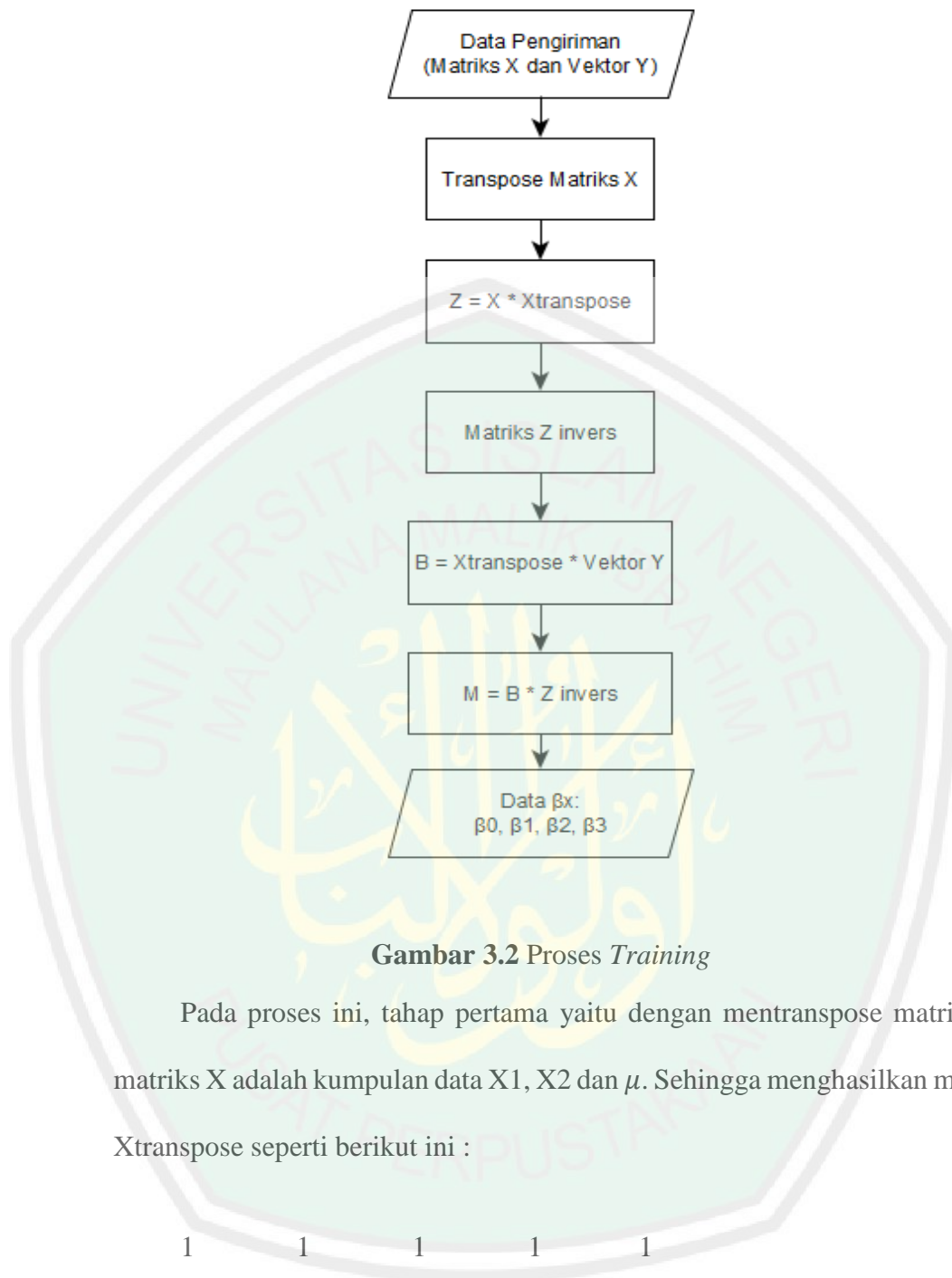
1	25	4000
1	25	4100
1	26	4200
1	26	4300
1	28	4400

Lakukan metode yang sama untuk membentuk matriks X dan vektor Y dari daerah lainnya (daerah 2, 3, 4, dan 5).

3.3.1.3 *Training Data dengan Metode Regresi Linear*

Tujuan dilakukannya *training data* adalah untuk mengetahui perbandingan nilai Y awal dan Y akhir yang merupakan nilai prediksi waktu pengiriman. Dengan cara ini pula didapat hasil β_0 dan β_x (β_1, β_2) yang akan disubstitusikan ke dalam persamaan regresi linear, yang mana persamaan ini nantinya akan digunakan pada data *testing* untuk memprediksi nilai Y saat ada data x baru yang dimasukkan.

Proses data *training* ditunjukkan oleh Gambar berikut:



Gambar 3.2 Proses *Training*

Pada proses ini, tahap pertama yaitu dengan mentranspose matriks X , matriks X adalah kumpulan data X_1 , X_2 dan μ . Sehingga menghasilkan matriks $X^{\text{transpose}}$ seperti berikut ini :

1	1	1	1	1
25	25	26	26	28
4000	4100	4200	4300	4400

Kemudian mengalikan matriks $X^{\text{transpose}}$ dengan dengan matriks X awal, sehingga menghasilkan matriks Z sebagai berikut,

5	130	21000
130	3386	546700
21000	546700	88300000

Lalu matriks Z diinverskan, menghasilkan matriks Z^{-1} ,

187,1091	3,090909	-0,06363636
3,090909	0,909091	-0,00636364
-0,06364	-0,00636	5,45455E-05

Selanjutnya mencari nilai matriks B yang merupakan hasil dari perkalian matriks X^T dengan vektor Y ,

320
8334
1346000

Terakhir yaitu mengalikan matriks Z^{-1} dengan matriks B , sehingga didapat sebuah hasil dalam bentuk vektor M yang merupakan nilai βx , sebagai berikut:

Tabel 3.7 Nilai Konstanta dan koefisien variabel x

βx	Nilai
β_0	-20
β_1	1,81899E-12
β_2	0,02

Dengan menggunakan metode *training* data yang sama, maka akan didapatkan nilai koefisien variabel X terhadap nilai Y daerah lain. Koefisien variabel X ini kemudian disubstitusikan ke dalam model persamaan linear untuk mengetahui

hasil persamaan regresi pada setiap daerah. Berikut adalah rincian hubungan nilai variabel X terhadap masing-masing nilai Y (Y_1, Y_2, \dots, Y_5).

Tabel 3.8 Koefisien Regresi Tiap Daerah

	X1	X2	C (β_0)
Y1	0.433803	0.0174946	-20.872
Y2	2.32826	0.00591749	-21.2943
Y3	0.460881	0.0172226	-17.8551
Y4	2.34397	0.00571084	-23.1955
Y5	0.385954	0.0179091	-25.7009

3.3.1.4 Persamaan Regresi Linear

Dari hasil data *training* pada Tabel 3.8 maka didapatkan persamaan linear :

$$Y_k = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$Y_1 = -20.872 + 0.433803X_1 + 0.0174946X_2$$

$$Y_2 = -21.2943 + 2.32826X_1 + 0.00591749X_2$$

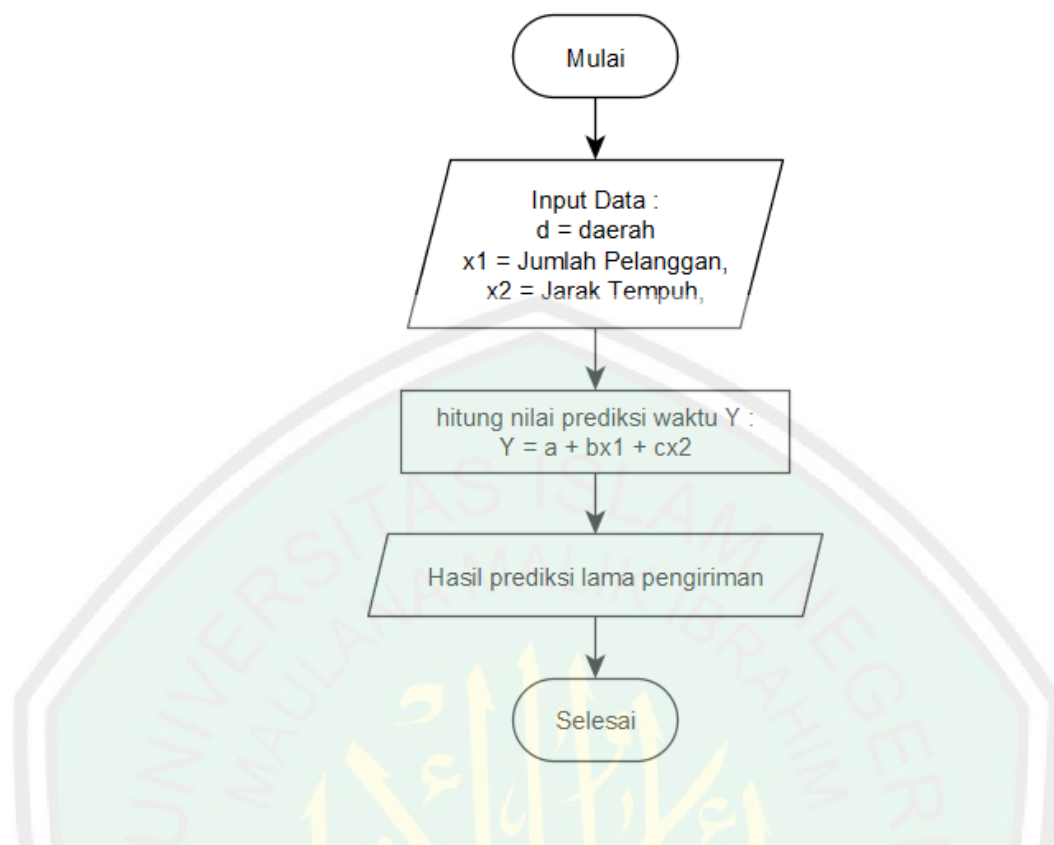
$$Y_3 = -17.8551 + 0.460881 X_1 + 0.0172226X_2$$

$$Y_4 = -23.1955 + 2.34397X_1 + 0.00571084X_2$$

$$Y_5 = -25.7009 + 0.385954X_1 + 0.0179091X_2$$

3.3.1.5 Testing Data

Persamaan linear yang telah diperoleh sebelumnya digunakan untuk memprediksi lama waktu pengiriman pada setiap data baru (*testing*) yang diinputkan ke sistem.



Gambar 3.3 Tahapan regresi linear data *testing*

3.3.1.6 Hasil Perbandingan Y awal dengan Y hasil *training*

Berikut adalah hasil perbandingan nilai Y awal dengan Y hasil proses *training* (nilai Y berdasarkan hasil penghitungan persamaan linear) :

Tabel 3.9 Perbandingan nilai Y awal dengan Y *training*

X1	X2	Y0	Y1	ϵ
25	4000	60	60	0
25	4100	62	62	0
26	4200	64	64	0
26	4300	66	66	0

BAB 4

HASIL DAN PEMBAHASAN

Dalam bab ini dibahas mengenai implementasi dari hasil perancangan sistem yang telah dirancang pada bahasan sebelumnya. Implementasi ini meliputi implementasi perancangan antarmuka (*interface*), implementasi sistem, uji statistik, *testing* dan analisa.

4.1 Implementasi Antarmuka/*Interface*

Semua bagian sistem ini terpampang dalam satu halaman ringkas yang terbagi menjadi beberapa blok program yang digunakan untuk memproses data.

Gambar 4.1 Halaman Utama Sebelum Data Diproses

4.1.1 Tampilan Load Data

Pada bagian ini berfungsi untuk melakukan proses dan menampilkan data hasil *preprocessing* yang akan diolah dalam penelitian ini. Data ini terdiri dari data jumlah pelanggan (X1), total jarak tempuh (X2), dan waktu pengiriman (Y) yang dikelompokkan berdasarkan daerah atau area pengiriman.

The screenshot shows a software application window titled "prediksi". On the left, there is a table of training data with columns X1, X2, and Y. Below the table is a legend explaining the variables: X1 = Jumlah Pe pelanggan, X2 = Total Jarak Tempuh (meter), and Y = Total Waktu Tempuh (menit). The main area of the application is titled "Training Data" and contains a "Training" button. Below this, there are five sections representing different regions: Daerah 1 (Soekarno-Hatta), Daerah 2 (Dinoyo), Daerah 3 (Tlogomas), Daerah 4 (Sumbersari), and Daerah 5 (BCT). Each region section has three input fields labeled B0, B1, and B2. On the right side, there is a "Prediksi" section with a dropdown menu for "Daerah Pengiriman", input fields for "Jumlah Pelanggan" and "Estimasi Jarak", and a "Hitung Waktu Tempuh" button.

	X1	X2	Y
1	25	4000	60
2	25	4100	62
3	26	4200	64
4	26	4300	66
5	28	4400	68
6	30	4500	70
7	30	4600	72
8	30	4700	74
9	28	4800	76
10	30	4900	78
11	29	5000	80
12	31	5100	84
13	30	5200	84
14	31	5300	86
15	35	5400	88
16	34	5500	88
17	34	5600	90
18	38	5700	96
19	36	5800	96
20	40	5800	101
21	30	4500	70
22	28	4800	76

Gambar 4.2 Tampilan Load Data

4.1.2 Tampilan Data Training

Halaman *data training* menampilkan data yang diolah dalam proses *training*. Data hasil *training* yang ditampilkan berupa nilai B0, B1, dan B2, serta visualisasi bentuk persamaan regresi yang akan digunakan untuk proses *testing*.

Persamaan yang dihasilkan dari proses *training* ini akan digunakan sebagai formula untuk memprediksi lama waktu pengiriman di bagian halaman *input* prediksi.

The screenshot shows a software interface for predicting delivery time. On the left, a table displays training data for 'Soekarno-Hatta' with columns X1, X2, and Y. The main area shows five regression equations for different regions, each with its own B0, B1, and B2 coefficients. The prediction section on the right includes a dropdown menu for 'Daerah Pengiriman', input fields for 'Jumlah Pelanggan' and 'Estimasi Jarak', and a 'Hitung Waktu Tempuh' button.

	X1	X2	Y
1	25	4000	60
2	25	4100	62
3	26	4200	64
4	26	4300	66
5	28	4400	68
6	30	4500	70
7	30	4600	72
8	30	4700	74
9	28	4800	76
10	30	4900	78
11	29	5000	80
12	31	5100	84
13	30	5200	84
14	31	5300	86
15	35	5400	88
16	34	5500	88
17	34	5600	90
18	38	5700	96
19	36	5800	96
20	40	5800	101
21	30	4500	70
22	28	4800	76

Daerah 1 (Soekarno-Hatta)
 $B_0 = -20.872$, $B_1 = 0.433803$, $B_2 = 0.0174946$
 $Y = -20.872 + 0.433803X_1 + 0.0174946X_2$

Daerah 2 (Dinoyo)
 $B_0 = -21.2943$, $B_1 = 2.32826$, $B_2 = 0.00591749$
 $Y = -21.2943 + 2.32826X_1 + 0.00591749X_2$

Daerah 3 (Tlogomas)
 $B_0 = -17.8551$, $B_1 = 0.460881$, $B_2 = 0.0172226$
 $Y = -17.8551 + 0.460881X_1 + 0.0172226X_2$

Daerah 4 (Sumbersari)
 $B_0 = -23.1955$, $B_1 = 2.34397$, $B_2 = 0.00571084$
 $Y = -23.1955 + 2.34397X_1 + 0.00571084X_2$

Daerah 5 (BCT)
 $B_0 = -25.7009$, $B_1 = 0.385954$, $B_2 = 0.0179091$
 $Y = -25.7009 + 0.385954X_1 + 0.0179091X_2$

Keterangan
 X_1 = Jumlah Peelanggan
 X_2 = Total Jarak Tempuh (meter)
 Y = Total Waktu Tempuh (menit)

Gambar 4.3 Tampilan *Data Training*

4.1.3 Tampilan *Input* untuk Memprediksi Lama Waktu Pengiriman

Blok tampilan ini digunakan untuk menerima inputan data baru yang nanti dapat menghasilkan prediksi lama waktu pengiriman yang dibutuhkan berdasarkan parameter-parameter yang telah ditentukan.

Prediksi

Soekarno-Hatta

Jumlah Pelanggan 40

Estimasi Jarak 7000

118.942

Hitung Waktu Tempuh

Gambar 4.4 Tampilan prediksi lama waktu pengiriman

Bagian *input* ini akan berfungsi atau aktif setelah sistem dilakukan proses *training* terlebih dahulu, karena proses prediksi ini membutuhkan hasil atau keluaran dari proses *training* untuk bisa mengestimasi lama waktu pengiriman.

Untuk melakukan prediksi terlebih dahulu diharuskan memilih daerah tujuan pengiriman, kemudian memasukkan data jumlah pelanggan dan estimasi jarak tempuh yang akan dilalui oleh kurir. Kemudian sistem akan menampilkan estimasi waktu yang dibutuhkan oleh kurir/pengirim untuk menyelesaikan proses pengiriman *catering* kepada seluruh pelanggannya.

Variabel masukan (*input*) berupa jumlah pelanggan, jarak tempuh dalam satuan meter, serta variabel *output* (keluaran) berupa waktu yang menggunakan satuan menit.

4.2 Implementasi Sistem

Sistem ini dibangun menggunakan bahasa pemrograman Matlab. Matlab dipilih penulis karena dinilai sangat cocok untuk mengolah data numerik yang

banyak menggunakan kaidah atau rumusan matematis dalam memecahkan suatu permasalahan.

4.2.1 Proses Load Data

Sistem ini membutuhkan data untuk menganalisis lama waktu pengiriman, data ini berupa data historis yang telah diolah sedemikian rupa sehingga memenuhi syarat untuk dilakukan proses regresi. Berikut adalah proses untuk menampilkan data pengiriman.

Tabel 4.1 Kode program proses load data

No	Source Code	Deskripsi
1	<pre>[num,txt,x] = xlsread('data.xlsx');</pre>	Kode ini berfungsi untuk me-load data. data tersebut merupakan data Microsoft excel dengan nama data.xlsx. dalam matlab memanggil data excel cukup dengan menggunakan function "xlsread".
2	<pre>set(handles.tbl_data,'data',x);</pre>	Bagian ini berfungsi untuk menampilkan data yang telah diload ke dalam tabel matlab di

		bagian <i>interface</i> . Nama variabelnya adalah <i>tbl_data</i> .
--	--	---

4.2.2 Proses *Training Data*

Training data dibagi menjadi lima kelompok, yaitu berdasarkan daerah pengiriman yang menjadi basis pelanggan dari Home Catering Malang, di antaranya daerah Soekarno-Hatta, Dinoyo, Tlogomas, Sumbersari dan Perumahan Bukit Cemara Tidar (BCT). Masing-masing daerah memiliki data sebanyak 25 *record* data untuk dilakukan proses *training*. Sehingga total data yang digunakan untuk proses *training* yaitu sebanyak 125 data.

Data tersebut diolah menggunakan metode regresi linear yang hasilnya berupa koefisien dari masing-masing variabel yang nantinya akan dimasukkan ke dalam persamaan linear. Secara total sistem ini akan memiliki lima macam persamaan linear yang mengacu pada setiap daerah pengiriman yang menjadi basis pelanggan Home Catering Malang. Kelima persamaan tersebut harus melalui uji asumsi statistik terlebih dahulu untuk memastikan agar hasil prediksi yang dilakukan oleh persamaan tersebut memiliki akurasi yang tinggi sehingga dapat meningkatkan kredibilitas sistem/aplikasi. Lebih lanjut tentang pengujian data dan persamaan regresi linear ini akan dibahas di bagian pengujian sistem.

Berikut adalah proses *training* untuk mendapatkan persamaan regresi linear berdasarkan data yang *ditraining*.

Tabel 4.2 Training data daerah 1 (Soekarno-Hatta)

No	Source Code	Deskripsi
1	<pre> %% %% data_training = xlsread('datalatih.xlsx','suhat' , 'A1:C26'); Y = xlsread('datalatih.xlsx','suhat' , 'D1:D26'); </pre>	<p>Tahap pertama yaitu memanggil datanya terlebih dahulu, dalam hal ini yaitu data <i>excel</i> dengan nama <i>datalatih.xlsx</i> pada <i>sheet</i> (halaman) <i>suhat</i> dengan <i>cell data</i> A1:C26 dan D1:D26 pada <i>cell Microsoft excel</i>.</p>
2	<pre> matriks_A = data_training; matriks_ATrans = transpose(matriks_A); matriks_B = matriks_ATrans * matriks_A; matriks_BInverse = inv(matriks_B); matriks_C = matriks_ATrans * Y; matriks_D = matriks_BInverse * matriks_C; Bb0 = matriks_D(1); Bb1 = matriks_D(2); Bb2 = matriks_D(3); </pre>	<p>Selanjutnya dilakukan metode regresi dengan operasi matriks hingga didapat nilai B0, B1 dan B2.</p>
3	<pre> set(handles.b0_suhat,'string',Bb 0); set(handles.b1_suhat,'string',Bb 1); set(handles.b2_suhat,'string',Bb 2); </pre>	<p>Nilai-nilai B0, B1, dan B2 kemudian disubstitusikan ke dalam persamaan linear</p>

<pre> data_b0 = get(handles.b0_suhat,'string'); data_b1 = get(handles.b1_suhat,'string'); data_b2 = get(handles.b2_suhat,'string'); PL = ['Y = ', data_b0 , ' + ' ,data_b1, 'X1 + ' ,data_b2, 'X2']; set(handles.pers_suhat,'string', PL); </pre>	regresi berganda sehingga membentuk persamaan Y pertama (persamaan linear untuk daerah pengiriman Soekarno-Hatta)
--	---

Selanjutnya untuk proses *training* data daerah pengiriman yang lain metodenya sama dengan proses *training* yang telah dibahas pada **Tabel 4.2**, hanya saja yang membedakan adalah data sumbernya yang diambil dari data pengiriman daerah yang bersangkutan.

Sehingga dengan metode *training* ini, sistem akan menghasilkan lima *output* atau model persamaan regresi linear. Lebih detailnya akan dijabarkan dalam Tabel 4.3 berikut ini:

Tabel 4.3 Daftar persamaan regresi yang terbentuk berdasarkan daerah setelah dilakukan proses *training*

No	Daerah	Koefisiaen X	Persamaan Regresi
1	Soekarno-Hatta	$B_0 = -20.872$ $B_1 = 0.433803$	$Y = -20.872 + 0.4338X_1 + 0.017495X_2$

		B2 = 0.0174946	
2	Dinoyo	B0 = -21.2943 B1 = 2.32826 B2 = 0.00591749	Y = -21.2943 + 2.32826X1 +0.00591749X2
3	Tlogomas	B0 = -17.8551 B1 = 0.460881 B2 = 0.0172226	Y = -17.8551 + 0.460881X1 +0.0172226X2
4	Sumbersari	B0 = -23.1955 B1 = 2.34397 B2 = 0.00571084	Y = -23.1955 + 2.34397X1 +0.00571084X2
5	BCT	B0 = -25.7009 B1 = 0.385954 B2 = 0.0179091	Y = -25.7009 + 0.385954X1 +0.0179091X2

4.2.3 Prediksi Lama Waktu Pengiriman

Halaman prediksi lama waktu pengiriman akan aktif setelah proses *training* selesai. Karena prediksi ini menggunakan persamaan regresi hasil *training* untuk

menghitung variabel *inputan* agar bisa diproses dan diprediksi total waktu pengirimannya. Berikut adalah penjabaran rincinya.

Tabel 4.4 Proses Prediksi lama waktu pengiriman

No	Source Code	Deskripsi
1	<pre>function hitung_waktu_Callback(hObject, eventdata, handles) pelanggan = str2double(get(handles.jml_pelang gan, 'string')); jrk = str2double(get(handles.jarak, 'str ing')); kirimkanke = get(handles.kirimkemana, 'Value');</pre>	<p>Kode ini berfungsi untuk menginisialisasi <i>input</i>. Setiap <i>inputan</i> akan disimpan ke dalam variabel. Supaya dapat dilakukan operasi penghitungan matematis maka data <i>input</i> dikonversi ke dalam tipe data <i>double</i> terlebih dahulu dari yang sebelumnya berupa tipe data <i>string</i>.</p>
2	<pre>switch kirimkanke case 2 %PL suhat b00 = str2double(get(handles.b0_suhat, ' string')); b11 = str2double(get(handles.b1_suhat, ' string'));</pre>	<p>Setelah diinisialisasi maka variabel bisa digunakan untuk proses penghitungan/prediksi. Kode di samping adalah proses substitusi</p>

<pre> b22 = str2double(get(handles.b2_suhat,' string')); wkt_tempuh = b00 + b11*pelanggan + b22*jrk; set(handles.waktu_tempuh,'string' ,wkt_tempuh); case 3 %PL dinoyo b00 = str2double(get(handles.b0_dinoyo, 'string')); b11 = str2double(get(handles.b1_dinoyo, 'string')); b22 = str2double(get(handles.b2_dinoyo, 'string')); wkt_tempuh = b00 + b11*pelanggan + b22*jrk; set(handles.waktu_tempuh,'string' ,wkt_tempuh); case 4 %PL tlogomas b00 = str2double(get(handles.b0_tlogoma s,'string')); b11 = str2double(get(handles.b1_tlogoma s,'string')); b22 = str2double(get(handles.b2_tlogoma s,'string')); wkt_tempuh = b00 + b11*pelanggan + b22*jrk; set(handles.waktu_tempuh,'string' ,wkt_tempuh); case 5 %PL sumbersari b00 = str2double(get(handles.b0_sumbers ari,'string')); </pre>	<p>variabel <i>input</i> ke dalam persamaan regresi linear. Seperti yang telah dijabarkan pada Tabel 4.3, Persamaan pertama adalah untuk menghitung lama waktu pengiriman kepada pelanggan yang berada di daerah Soekarno-Hatta, persamaan yang kedua untuk daerah Dinoyo, ketiga adalah Tlogomas, keempat adalah Sumbersari, dan yang terakhir untuk wilayah BCT.</p>
---	---

```

        b11 =
str2double(get(handles.b1_sumbers
ari,'string'));
        b22 =
str2double(get(handles.b2_sumbers
ari,'string'));
        wkt_tempuh = b00 +
b11*pelanggan + b22*jrk;

set(handles.waktu_tempuh,'string'
,wkt_tempuh);

        case 6
        %PL BCT
        b00 =
str2double(get(handles.b0_bct,'st
ring'));
        b11 =
str2double(get(handles.b1_bct,'st
ring'));
        b22 =
str2double(get(handles.b2_bct,'st
ring'));
        wkt_tempuh = b00 +
b11*pelanggan + b22*jrk;

set(handles.waktu_tempuh,'string'
,wkt_tempuh);
end

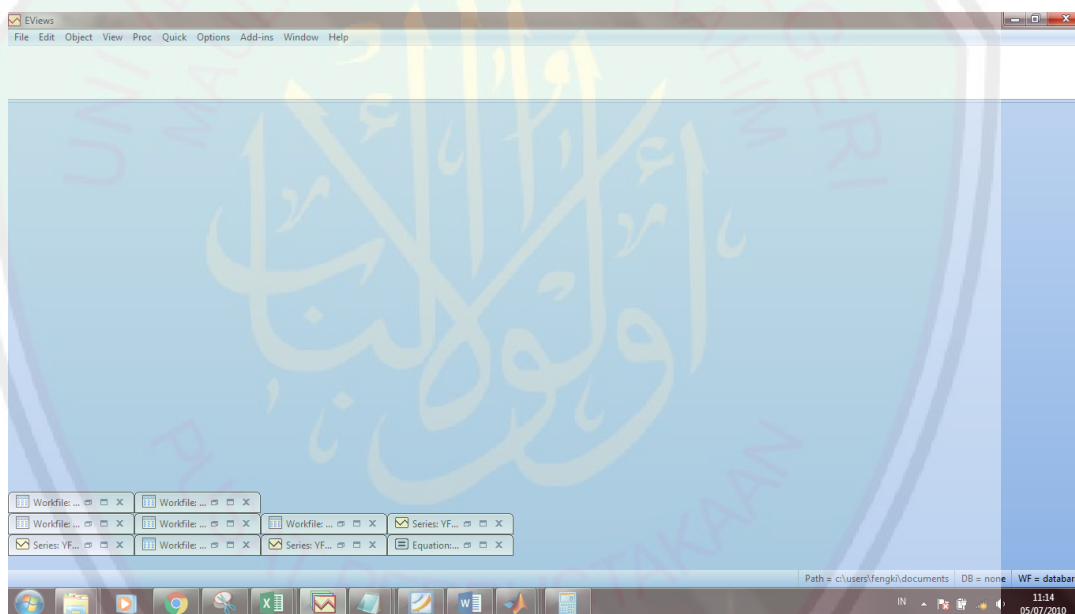
```

4.3 Uji Asumsi Statistik

Pengujian ini untuk memastikan bahwa persamaan regresi linear yang dihasilkan oleh sistem dapat memprediksi lama waktu pengiriman secara maksimal. Pengujian ini meliputi beberapa aspek uji asumsi klasik regresi linear, diantaranya yaitu multikolinieritas, linearitas, heteroskedastisitas, autokorelasi, dan normalitas. Namun penekanan utama akan lebih dititikberatkan pada nilai koefisien determinasi dari gabungan variabel X, semakin tinggi nilai koefisien determinasi, semakin kuat pula pengaruh dari variabel X dalam persamaan regresi terhadap variabel Y (waktu pengiriman). Persamaan regresi memiliki nilai koefisien determinasi yang bagus

apabila semakin mendekati 1, sebaliknya akan semakin tidak bagus pula apabila nilainya mendekati 0. Apabila koefisien determinasi mendekati atau bahkan sama dengan 0, maka persamaan regresi tersebut tidak bisa mewakili atau menggambarkan hubungan variabel X terhadap Y. Artinya persamaan linear yang dihasilkan tidak/kurang relevan dan tidak bisa dijadikan acuan untuk melakukan prediksi.

Agar hasil pengujiannya dapat lebih dipertanggungjawabkan, maka penulis menggunakan *software* atau aplikasi statistik untuk melakukan analisa pengujian ini. Aplikasi yang digunakan adalah *Eviews* versi 7.



Gambar 4.5 Halaman antar muka *EViews* 7

Berikut detail pengujian dari aplikasi *Eviews 7*

1. Pengujian pada daerah 1 (Soekarno-Hatta)

Tabel 4.5 Pengujian untuk daerah Soekarno-Hatta

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.433803	0.156861	2.765531	0.0113
X2	0.017495	0.001110	15.76692	0.0000
C	-20.87199	2.201620	-9.480288	0.0000
R-squared	0.989781	Mean dependent var	79.00000	
Adjusted R-squared	0.988852	S.D. dependent var	11.52533	
S.E. of regression	1.216899	Akaike info criterion	3.342655	
Sum squared resid	32.57855	Schwarz criterion	3.488920	
Log likelihood	-38.78319	Hannan-Quinn criter.	3.383223	
F-statistic	1065.414	Durbin-Watson stat	1.680031	
Prob(F-statistic)	0.000000			

Dari hasil pada **Tabel 4.5** di atas dapat disimpulkan bahwa koefisien determinasi variabel X terhadap nilai Y sangat tinggi (hampir mendekati 1), yaitu *R-squared* sebesar 0.989781 atau sekitar 98.99% dan *Adjusted R-squared* sebesar 98.99% pula. Hal ini menunjukkan variabel lain yang berpengaruh terhadap nilai Y selain variabel X1 (jumlah pelanggan) dan X2 (total jarak tempuh) sangat **minim**, sehingga persamaan ini dapat menggambarkan hubungan variabel-variabel X terhadap Y (waktu pengiriman) secara kuat.

Beberapa pengujian lain terhadap analisis regresi yang dilakukan melalui aplikasi *Eviews 7* hasilnya sebagai berikut:

Tabel 4.6 Hasil uji asumsi klasik

Nama Uji Asumsi Klasik	Metode Pengujian	Hasil	Kesimpulan
Multikolinearitas	Uji <i>VIF</i>	<i>VIF</i> X1 dan X2 = 6,336562	Tidak terjadi multikolinearitas karena nilai <i>VIF</i> tidak lebih dari 10
Autokorelasi	<i>LM Test</i>	<i>Prob. F-Statistic</i> = 0,4606	Tidak terjadi autokorelasi karena nilai <i>Prob. F-Stat</i> > 0,05
Linearitas	<i>Ramsey</i> <i>RESET Test</i>	<i>Prob. F-Statistic</i> = 0,6828	Memenuhi linearitas karena <i>Prob. F-Stat</i> > 0,05
Normalitas	<i>Histogram</i> <i>Normality Test</i>	<i>Prob. JB</i> = 0,4902	Residual terdistribusi normal karena <i>Prob. JB</i> > 0,05

Dengan melihat hasil analisis uji asumsi klasik di atas maka dapat disimpulkan bahwa persamaan regresi untuk daerah 1 (Soekarno-Hatta) layak digunakan untuk memodelkan hubungan variabel X terhadap Y karena sangat minim terjadinya bias.

2. Pengujian terhadap daerah yang lain

Dengan menggunakan metode yang sama dengan pengujian pada **Tabel 4.5** dan **Tabel 4.6** di atas, berikut hasil pengujian dan analisis pada persamaan regresi linear untuk empat daerah yang lain:

Tabel 4.7 Uji asumsi klasik untuk masing-masing daerah

Daerah	Pengujian	Hasil	Kesimpulan
Dinoyo	Multikolinearitas	<i>VIF</i> X1 dan X2 = 5,998	Tidak terjadi multikolinearitas karena nilai <i>VIF</i> tidak lebih dari 10
	Autokorelasi	<i>Prob. F-Statistic</i> = 0,1609	Tidak terjadi autokorelasi karena nilai <i>Prob. F-Stat</i> > 0,05
	Linearitas	<i>Prob. F-Statistic</i> = 0,5411	Memenuhi linearitas karena <i>Prob. F-Stat</i> > 0,05
	Normalitas	<i>Prob. JB</i> = 0,056	Residual terdistribusi normal karena <i>Prob. JB</i> > 0,05
Tlogomas	Multikolinearitas	<i>VIF</i> X1 dan X2 = 6,2678	Tidak terjadi multikolinearitas karena nilai <i>VIF</i> tidak lebih dari 10
	Autokorelasi	<i>Prob. F-Statistic</i> = 0,057	Tidak terjadi autokorelasi karena nilai <i>Prob. F-Stat</i> > 0,05
	Linearitas	<i>Prob. F-Statistic</i> = 0,6908	Memenuhi linearitas karena <i>Prob. F-Stat</i> > 0,05
	Normalitas	<i>Prob. JB</i> = 0,8543	Residual terdistribusi normal karena <i>Prob. JB</i> > 0,05
Sumber-Sari	Multikolinearitas	<i>VIF</i> X1 dan X2 = 6,016	Tidak terjadi multikolinearitas karena nilai <i>VIF</i> tidak lebih dari 10
	Autokorelasi	<i>Prob. F-Statistic</i> = 0,1431	Tidak terjadi autokorelasi karena nilai <i>Prob. F-Stat</i> > 0,05

	Linearitas	<i>Prob. F-Statistic</i> = 0,5383	Memenuhi linearitas karena <i>Prob. F-Stat</i> > 0,05
	Normalitas	<i>Prob. JB</i> = 0,06	Residual terdistribusi normal karena <i>Prob. JB</i> > 0,05
BCT	Multikolinearitas	<i>VIF</i> X1 dan X2 = 6,450	Tidak terjadi multikolinearitas karena nilai <i>VIF</i> tidak lebih dari 10
	Autokorelasi	<i>Prob. F-Statistic</i> = 0,059	Tidak terjadi autokorelasi karena nilai <i>Prob. F-Stat</i> > 0,05
	Linearitas	<i>Prob. F-Statistic</i> = 0,7576	Memenuhi linearitas karena <i>Prob. F-Stat</i> > 0,05
	Normalitas	<i>Prob. JB</i> = 0,249	Residual terdistribusi normal karena <i>Prob. JB</i> > 0,05

Berdasarkan hasil analisa pada **Tabel 4.7** di atas, dapat disimpulkan bahwa setiap persamaan regresi linear memenuhi syarat untuk dikatakan sebagai model terbaik untuk memprediksi lama waktu pengiriman *catering*. Sedangkan koefisien determinasi dapat dilihat pada tabel berikut:

Tabel 4.8 Nilai koefisien determinasi pada setiap persamaan

Persamaan	<i>R-Squared</i>	<i>Adjusted R-Squared</i>
Persamaan 2 (Dinoyo) = -21,2943+2,32826X1+0,00591749X2	0,956753	0,952822
Persamaan 3 (Tlogomas) = -17.8551 +0.460881X1+0.0172226X2	0,992183	0,991472

Persamaan 4 (Sumbersari) = -23.1955 +2.34397X1+0.00571084X2	0,9564	0,9524
Persamaan 5 (BCT) = -25,7009 +0,385954X1+0,0179091X2	0,989	0,9875

Dari **Tabel 4.8** di atas maka dapat dilihat bahwa nilai koefisien determinasi pada masing-masing model regresi di atas 90%, artinya model ini sangat bisa menggambarkan hubungan variabel X terhadap Y dengan tingkat pengaruh yang sangat besar.

4.4 Uji Akurasi Hasil Prediksi

Pada subbahasan uji akurasi ini, sistem akan diuji kemampuan prediksinya dengan menggunakan data *testing*. Masing-masing persamaan akan diuji menggunakan 15 *records* data. Sehingga total data testing yang digunakan pada pengujian ini berjumlah 75 *records* data.

Dari pengujian ini akan dilihat seberapa besar tingkat kemiripan nilai Y prediksi dengan nilai Y aktual (Y yang sebenarnya dari data uji yang sudah terhimpun). Untuk mengetahuinya, penulis menggunakan metode *Mean Absolute Deviation* (MAD) dan *Mean Absolute Percentage Error* (MAPE).

1. Pengujian dengan Persamaan 1 (Soekarno-Hatta)

Persamaan regresi yang akan digunakan untuk daerah pengiriman Soekarno-Hatta adalah $Y = -20.872 + 0.433803X1 + 0.0174946X2$. Berikut adalah hasil *testing* data dengan menggunakan persamaan tersebut,

Tabel 4.9 Pengujian dengan persamaan 1

No.	X1	X2	Y	Yt	Abs(error)	%error
1	26	4100	64	62,13474	1,865262	2,914472
2	26	4200	66	63,8842	2,115802	3,205761
3	27	4300	68	66,06746	1,932539	2,841969
4	27	4400	70	67,81692	2,183079	3,118684
5	29	4500	72	70,43399	1,566013	2,175018
6	31	4600	74	73,05105	0,948947	1,282361
7	31	4700	76	74,80051	1,199487	1,578272
8	31	4800	78	76,54997	1,450027	1,859009
9	29	4900	80	77,43183	2,568173	3,210216
10	31	5000	82	80,04889	1,951107	2,379399
11	30	5100	84	81,36455	2,63545	3,13744
12	32	5200	88	83,98162	4,018384	4,566345
13	31	5300	88	85,29727	2,702727	3,071281
14	32	5400	90	87,48054	2,519464	2,799404
15	36	5500	92	90,96521	1,034792	1,124774
<i>Total error</i>					30,69125	39,26441
<i>Mean Absolute Deviation (MAD)</i>					30,69125/15 = 2,046084	
<i>Mean Absolute Percentage Error (MAPE)</i>					39,26441/15 = 2,617627%	

Dari data *testing* di atas, X1 dan X2 merupakan variabel *input*, sedangkan Y adalah data Y sebenarnya yang menunjukkan lama waktu pengiriman, Yt merupakan nilai Y hasil *testing* menggunakan persamaan 1.

Kemudian akan dihitung akurasiya menggunakan metode MAD dan MAPE. Sebagaimana yang tertulis dalam tabel di atas, persamaan ini mampu memprediksi nilai Y dengan taraf *error* MAD = 2,046 dan persentase *error*-nya jika dihitung dengan metode MAPE menghasilkan 2,618% tingkat *error*.

2. Pengujian dengan Persamaan 2 (Dinoyo)

Persamaan regresi yang akan digunakan untuk daerah pengiriman Dinoyo adalah $Y = -21.2943 + 2.32826X_1 + 0.00591749X_2$. Berikut adalah hasil *testing* data dengan menggunakan persamaan tersebut,

Tabel 4.10 Pengujian dengan persamaan 2

No.	X1	X2	Y	Yt	Abs(error)	%error
1	26	4150	65	63,79804	1,201956	1,849164
2	26	4350	69	64,98154	4,018458	5,823853
3	28	4450	73	70,22981	2,770189	3,79478
4	28	4400	70	69,93394	0,066064	0,094377
5	25	4100	63	61,17391	1,826091	2,898557
6	29	4450	75	72,55807	2,441929	3,255906
7	25	3950	60	60,28629	0,286286	0,477143
8	25	3950	60	60,28629	0,286286	0,477143
9	21	3650	50	49,198	0,802001	1,604003
10	22	3750	52	52,11801	0,118008	0,226938
11	19	3550	47	43,94973	3,050271	6,489937
12	19	3550	47	43,94973	3,050271	6,489937
13	29	4200	67	71,0787	4,078698	6,087609

14	29	4200	67	71,0787	4,078698	6,087609
15	27	3900	70	64,64693	5,353069	7,647241
Total error					33,42827	53,3042
<i>Mean Absolute Deviation (MAD)</i>					33,42827/15 = 2,228552	
<i>Mean Absolute Percentage Error (MAPE)</i>					53,3042/15 = 3,553613%	

Dari data *testing* di atas, X1 dan X2 merupakan variabel *input*, sedangkan Y adalah data Y sebenarnya yang menunjukkan lama waktu pengiriman, Yt merupakan nilai Y hasil *testing* menggunakan persamaan 2. Hasil uji akurasi menunjukkan bahwa model persamaan tersebut mampu memprediksi lama waktu pengiriman Y dengan tingkat atau persentase kesalahan (*error*) sebesar 3,55% berdasarkan metode pengujian *Mean Absolute Percentage Error (MAPE)* dan nilai MAD sebesar 2.23.

3. Pengujian dengan Persamaan 3 (Daerah Tlogomas)

Persamaan regresi yang akan digunakan untuk daerah pengiriman Tlogomas adalah $Y = -17.8551 + 0.460881X1 + 0.0172226X2$. Berikut adalah hasil *testing* data dengan menggunakan persamaan tersebut:

Tabel 4.11 Pengujian dengan persamaan 3

No.	x1	x2	Y	Yt	Abs(error)	%error
1	29	4300	71	69,56763	1,432371	2,017424
2	29	4420	73	71,63434	1,365659	1,870766
3	30	4530	75	73,98971	1,010292	1,347056
4	30	4620	77	75,53974	1,460258	1,896439

5	32	4720	79	78,18376	0,816236	1,03321
6	32	4700	79	77,83931	1,160688	1,469225
7	34	4820	81	80,82779	0,172214	0,21261
8	34	4820	81	80,82779	0,172214	0,21261
9	34	4900	83	82,20559	0,794406	0,957116
10	34	5020	85	84,27231	0,727694	0,856111
11	32	5110	87	84,90058	2,099422	2,413129
12	32	5120	87	85,0728	1,927196	2,215168
13	34	5220	89	87,71683	1,283174	1,441769
14	33	5300	91	88,63375	2,366247	2,600271
15	35	5420	95	91,62223	3,377773	3,555551
Total error					20,16584	24,09845
<i>Mean Absolute Deviation (MAD)</i>					20,16584/15 = 1,34439	
<i>Mean Absolute Percentage Error (MAPE)</i>					24,09845/15 = 1,606564%	

Dari data *testing* di atas, X1 dan X2 merupakan variabel *input*, sedangkan Y adalah data Y sebenarnya yang menunjukkan lama waktu pengiriman, Yt merupakan nilai Y hasil *testing* menggunakan persamaan 3. Hasil uji akurasi menunjukkan bahwa model persamaan tersebut mampu memprediksi lama waktu pengiriman Y dengan tingkat atau persentase kesalahan (*error*) sebesar 1,61% berdasarkan metode pengujian *Mean Absolute Percentage Error (MAPE)* dan nilai MAD sebesar 1.34.

4. Pengujian dengan Persamaan 4 (Daerah Summersari)

Persamaan regresi yang akan digunakan untuk daerah pengiriman Tlogomas adalah $Y = -23.1955 + 2.34397X_1 + 0.00571084X_2$. Berikut adalah hasil *testing* data dengan menggunakan persamaan tersebut,

Tabel 4.12 Pengujian dengan persamaan 4

No.	x1	x2	Y	Yt	Abs(error)	%error
1	26	4000	60	60,59108	0,59108	0,985133
2	26	4150	64	61,44771	2,552294	3,987959
3	28	4320	68	67,10649	0,893511	1,313987
4	28	4250	65	66,70673	1,70673	2,625738
5	25	3950	58	57,96157	0,038432	0,066262
6	29	4300	70	69,33624	0,663758	0,948226
7	25	3800	55	57,10494	2,104942	3,827167
8	27	3750	65	61,50734	3,49266	5,373323
9	27	3750	65	61,50734	3,49266	5,373323
10	25	3800	55	57,10494	2,104942	3,827167
11	21	3500	45	46,01581	1,01581	2,257356
12	22	3600	47	48,93086	1,930864	4,108221
13	19	3400	42	40,75679	1,243214	2,960033
14	19	3400	42	40,75679	1,243214	2,960033
15	26	3600	60	58,30674	1,693256	2,822093
Total error					24,76737	43,43602
<i>Mean Absolute Deviation (MAD)</i>					24,76737/15 = 1,651158	

<i>Mean Absolute Percentage Error (MAPE)</i>	$43,43602/15 = 2,895735\%$
--	----------------------------

Hasil uji akurasi menunjukkan bahwa model persamaan pada daerah pengiriman Summersari mampu memprediksi lama waktu pengiriman Y dengan tingkat atau persentase kesalahan (*error*) sebesar 2.9% berdasarkan metode pengujian *Mean Absolute Percentage Error (MAPE)* dan nilai MAD sebesar 1.65.

5. Pengujian dengan Persamaan 5 (Daerah BCT)

Persamaan regresi yang akan digunakan untuk daerah pengiriman Tlogomas adalah $Y = -25.7009 + 0.385954X_1 + 0.0179091X_2$. Berikut adalah hasil *testing* data dengan menggunakan persamaan tersebut,

Tabel 4.13 Pengujian dengan persamaan 5

No.	x1	x2	Y	Yt	<i>Abs(error)</i>	<i>%error</i>
1	23	4700	67	67,34881	0,348812	0,520615
2	23	4800	69	69,13972	0,139722	0,202496
3	24	4900	71	71,31659	0,316586	0,445896
4	24	5000	73	73,1075	0,107496	0,147255
5	26	5100	75	75,67031	0,670314	0,893752
6	26	5100	75	75,67031	0,670314	0,893752
7	28	5200	77	78,23313	1,233132	1,60147
8	28	5200	77	78,23313	1,233132	1,60147
9	28	5300	79	80,02404	1,024042	1,296256
10	28	5400	81	81,81495	0,814952	1,006114
11	26	5500	83	82,83395	0,166046	0,200055

12	26	5500	83	82,83395	0,166046	0,200055
13	28	5600	85	85,39677	0,396772	0,466791
14	27	5700	87	86,80173	0,198272	0,227899
15	29	5800	91	89,36455	1,635454	1,797202
<i>Total error</i>					9,121092	11,50108
<i>Mean Absolute Deviation (MAD)</i>					9,121092/15 = 0,608073	
<i>Mean Absolute Percentage Error (MAPE)</i>					11,50108/15 = 0,766738%	

Hasil uji akurasi menunjukkan bahwa model persamaan pada daerah pengiriman BCT mampu memprediksi lama waktu pengiriman Y dengan tingkat atau persentase kesalahan (*error*) sebesar 0,77% berdasarkan metode pengujian *Mean Absolute Percentage Error* (MAPE) dan nilai MAD sebesar 0,61.

Dari hasil uji akurasi nilai Y prediksi pada masing-masing daerah di atas dapat dibandingkan hasilnya seperti pada **Tabel 4.14** di bawah ini :

Tabel 4.14 Perbandingan hasil uji akurasi

No	Persamaan Regresi	MAD	MAPE
1	Persamaan 1 (Soekarno Hatta)	2,046084	2,617627 %
2	Persamaan 2 (Dinoyo)	2,228552	3,553613 %
3	Persamaan 3 (Tlogomas)	1,34439	1,606564 %
4	Persamaan 4 (Sumbersari)	1,651158	2,895735 %
5	Persamaan 5 (BCT)	0,608073	0,766738 %
Rata-rata		1,575651	2,29%

Dari **Tabel 4.14** dapat diketahui bahwa kelima persamaan regresi yang digunakan dalam sistem ini memiliki kemampuan untuk memprediksi nilai Y dengan level *error* yang sangat minim. Dengan kata lain kelima persamaan tersebut memiliki akurasi yang tinggi. Berdasarkan data hasil pengujian (*testing*), persamaan dengan akurasi paling tinggi dimiliki oleh persamaan 5 (daerah BCT) karena memiliki level *error* paling kecil di antara persamaan yang lain (MAD = 0,61 dan MAPE = 0,77%). Lalu persamaan 2 (daerah Dinoyo) sebagai persamaan dengan akurasi prediksi paling rendah dibanding keempat persamaan lainnya (MAD = 2,23 dan MAPE = 3,55%). Jika dirata-rata, kelima model persamaan regresi ini dapat melakukan prediksi dengan level *error* rata-rata MAD sebesar 1,58 dan MAPE sebesar 2,29 %.

4.5 Integrasi dalam Islam

Penelitian ini erat kaitannya dengan ayat dalam Al-Quran yaitu surat Al-Insyirah. Sebagaimana yang berbunyi:

أَلَمْ نَشْرَحْ لَكَ صَدْرَكَ (1) وَوَضَعْنَا عَنكَ وِزْرَكَ (2) الَّذِي أَنْقَضَ ظَهْرَكَ (3)
وَرَفَعْنَا لَكَ ذِكْرَكَ (4) فَإِنَّ مَعَ الْعُسْرِ يُسْرًا (5) إِنَّ مَعَ الْعُسْرِ يُسْرًا (6) فَإِذَا
فَرَغْتَ فَاَنْصَبْ (7) وَإِلَى رَبِّكَ فَارْغَبْ (8)

Artinya :” Bukankah Kami telah melapangkan untukmu dadamu? dan Kami telah menghilangkan daripadamu bebanmu, yang memberatkan punggungmu? Dan Kami tinggikan bagimu sebutan (nama)mu. Karena sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain. dan hanya kepada Tuhanmulah hendaknya kamu berharap”.

Dalam Tafsir Ibnu Katsir dijelaskan bahwa maksud dari ayat yang berbunyi “*Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain*”, maksudnya adalah, jika seseorang telah selesai mengurus berbagai kepentingan dunia dan semua kesibukannya serta telah mengurus semua jaringannya, maka bersungguh-sungguhlah untuk menjalankan ibadah serta melangkahlah kepada –Nya dengan penuh semangat, dengan hati yang kosong lagi tulus, serta niat karena Allah. (Muhammad, 2004)

Ayat ini mengajarkan betapa kita harus bersungguh-sungguh dalam pekerjaan. Jika dikaitkan dengan pengiriman *catering*, maka akan sangat positif manfaatnya apabila selalu tepat waktu dalam menyelesaikan pengiriman kepada pelanggan. Waktu yang tersedia setelahnya bisa dipergunakan untuk segera menyelesaikan pekerjaan yang lain. Dengan begitu saat datang beribadah kepada Allah kita bisa khusyu’, dan orang yang kita layani melalui pekerjaan kita akan merasa puas dan semakin percaya, sehingga berpeluang besar untuk lebih meningkatkan bisnis dan meraih rejeki atas ijin Allah SWT.

BAB 5

KESIMPULAN

5.1 Kesimpulan

Metode regresi linear yang diterapkan dalam sistem ini dapat memprediksi lama waktu pengiriman *catering* kepada pelanggan dengan tingkat *error* yang sangat kecil. Dengan metode *Mean Absolute Percentage Error* (MAPE) didapatkan tingkat *error* rata-rata kelima persamaan regresi linear adalah 2,29 persen dan rata-rata *Mean Absolute Deviation* (MAD) sebesar 1,58. Akurasi tertinggi dimiliki oleh persamaan 5 (daerah pengiriman BCT) karena memiliki tingkat *error* paling kecil (MAD = 0,61 dan MAPE = 0,77%). Sedangkan akurasi paling rendah dimiliki oleh persamaan 2 (Dinoyo) karena memiliki tingkat *error* yang lebih besar dibanding keempat persamaan lainnya (MAD = 2,23 dan MAPE = 3,55%).

Nilai MAD dan MAPE berbanding lurus. Semakin besar nilai MAD, maka semakin besar pula nilai MAPE-nya. Namun keduanya berbanding terbalik dengan akurasi. Semakin besar nilai MAD dan MAPE, maka akurasi prediksinya semakin rendah. Sebaliknya, akurasi hasil prediksi sebuah persamaan linear akan semakin tinggi jika nilai MAD dan MAPE semakin kecil.

5.2 Saran

Untuk penelitian selanjutnya diharapkan dapat melengkapi dengan parameter/variabel lain agar bisa meningkatkan akurasi dari prediksi yang diberikan oleh sistem ini, atau setidaknya menjadi pembanding apabila ada variabel baru yang dapat mempengaruhi nilai Y. Serta dapat mengembangkan ke berbagai platform sehingga penggunaannya bisa lebih fleksibel.

DAFTAR PUSTAKA

- Algifari. (2000). *Analisis Regresi, Teori, Kasus & Solusi*. Yogyakarta: BPFE UGM.
- Bowerman, B. d. (1993). *Forecasting and Time Series: An Applied Approach, Edisi Ketiga*. California: Duxbury Press.
- Disa, S. (2011). Implementasi Metode Regresi Linear Dalam Pembuatan Aplikasi Simulasi. *Sistem Informasi STMIK AKBA Makassar*.
- Douglas C Montgomery & Team. (2013). *Applied Statistics and Probability for Engineers Edisi ke 6*. New Jersey: John Willey & Sons, Inc.
- Ghozali, I. (2005). *Aplikasi Analisis Multivariate dengan SPSS*. Semarang: UNDIP.
- Koro, P. K. (2017). Prediksi Tingkat Produksi Kopi Menggunakan Regresi Linear. *Flash, vol 3*.
- Kuncoro, M. (2013). *Metode Riset Untuk Bisnis & Ekonomi, Edisi Keempat*. Jakarta: Erlangga.
- Kusrini, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Lungan, R. (2006). *Aplikasi Statistika dan Hitung Peluang*. Bandung: Graha Ilmu.
- McLeod, R. (1995). *Sistem Informasi Manajemen*. Klaten : PT. Tema Baru .
- Muhammad, D. A. (2004). *Tafsir Ibnu Katsir Jilid 1*. Jakarta: Pustaka Imam Asy-Syafi'i.
- R Adisetiawan, Y. S. (2011). Analisa Pengaruh Variabel-variabel Fundamental terhadap Return Saham LQ45. *11(2)*.
- Robert Kurniawan, B. Y. (2016). *Analisis Regresi, Dasar dan Penerapannya dengan R*. Jakarta: Kencana.
- Susanto, S. d. (2010). *Pengantar Data* . Yogyakarta: Penerbit Andi.
- Turban, E. A. (2005). *Decision Support System and Intelligence System (Versi Bahasa Indonesia), Edisi Ke-7*. Yogyakarta: ANDI Offset.