BAB II
KAJIAN TEORI

2.1 Pengangguran

2.1.1 Pengertian Pengangguran

Menurut Badan Pusat Statistik (BPS) dalam indikator ketenagakerjaan, pengangguran adalah penduduk yang tidak bekerja namun sedang mencari pekerjaan atau sedang mempersiapkan suatu usaha baru atau penduduk yang tidak mencari pekerjaan karena sudah diterima bekerja tetapi belum mulai bekerja. Menurut Hasyim (2017:197) pengangguran merupakan masalah strategi dalam perekonomian secara makro, karena berpengaruh langsung kepada standar kehidupan dan tekanan psikologis masyarakat. Sedangkan menurut Nanga (2001:253) pengangguran adalah suatu keadaan dimana seseorang yang tergolong dalam kategori angkatan kerja tidak memiliki pekerjaan dan secara aktif tidak sedang mencari pekerjaan.

Pengangguran atau bisa disebut juga tunakarya adalah istilah untuk orang yang tidak bekerja sama sekali, atau sedang mencari pekerjaan. Pengangguran umumnya disebabkan karena jumlah angkatan kerja atau para pencari kerja tidak sebanding dengan jumlah lapangan kerja yang tersedia.Pengangguran merupakan masalah dalam perekonomian, karena dengan adanya pengangguran produktivitas dan pendapatan masyarakat berkurang sehingga menyebabkan timbulnya kemiskinan dan masalah-masalah sosial lainnya.

2.1.2 Jenis Pengangguran

Menurut Ritonga dan Firdaus (2007:8-11) jenis-jenis pengangguran dikelompokkan menjadi dua, yaitu:

a. Jenis pengangguran menurut lama waktu kerja

1) Pengangguran Terbuka

Pengangguran terbuka adalah tenaga kerja yang tidak mempunyai pekerjaan dan telah berusaha mencari pekerjaan secara maksimal, sementara lapangan kerja yang tersedia tidak cocok dengan latar belakang pendidikannya atau malas mencari pekerjaan.

2) Setengah Menganggur

Setengah menganggur adalah tenaga kerja yang tidak bekerja secara optimal karena ketiadaan lapangan kerja atau pekerjaan dan bekerja kurang dari 35 jam selama seminggu.

3) Pengangguran Terselubung

Pengangguran terselubung adalah tenaga kerja yang tidak bekerja secara optimal karena tidak memperoleh pekerjaan yang sesuai dengan bakat dan kemampuannya.

b. Jenis pengangguran menurut penyebab terjadinya

1) Pengangguran Struktural

Pengangguran struktural disebabkan oleh ketidakcocokan antara keterampilan tenaga kerja yang dibutuhkan dengan tenaga kerja yang tersedia berupa perubahan struktur permintaan penawaran dalam jangka panjang sebagai dampak kemajuan tekonologi, perubahan selera, dan persaingan antar perusahaan.

2) Pengangguran Siklikal

Pengangguran Siklikal berkaitan dengan naik turunnya aktivitas atau keadaan perekonomian suatu negara yang mengalami masa pertumbuhan atau mengalami penurunan bahan depresi.

3) Pengangguran Musiman

Pengangguran musiman disebabkan oleh perubahan permintaan terhadap tenaga kerja yang sifatnya berkala.Pengangguran ini biasanya terjadi pada tenaga kerja paruh waktu.

4) Pengangguran Friksional

Pengangguran friksional disebabkan oleh pergantian pekerjaan atau pergeseran tenaga kerja atau berpindah dari jenis pekerjaan tertentu ke jenis pekerjaan lain.

2.2 Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka

Tingkat pengangguran terbuka dapat dipengaruhi oleh beberapa faktor. Berikut ini merupakan faktor yang mempengaruhi tingkat pengangguran terbuka:

2.2.1 Pertumbuhan Ekonomi

Menurut Sadono Sukirno (2008) pertumbuhan ekonomi diartikan sebagai perkembangan kegiatan dalam perekonomian yang menyebabkan barang dan jasa yang diproduksi dalam masyarakat bertambah dan kemakmuran masyarakat meningkat. Dengan demikian untuk menentukan laju pertumbuhan ekonomi yang perlu dicapai perlu dihitung adalah pendapatan nasional rill menurut harga tetap yaitu harga berlaku ditahun dasar yang dipilih. Sehingga dapat dikatakan pertumbuhan ekonomi mangukur prestasi dari perkembangan perekonomian suatu negara. Pertumbuhan ekonomi adalah sebagian dari perkembangan kesejahteraan masyarakat yang diukur dengan besarnya pertumbuhan domestik regional bruto per kapita (PDRB per kapita).

PDRB menurut Badan Pusat Statistik (2018) yaitu jumlah nilai tambah yang dihasilkan untuk seluruh wilayah usaha dalam suatu wilayah atau merupakan jumlah seluruh nilai barang dan jasa akhir yang dihasilkan seluruh unit ekonomi di suatu wilayah. Menurut departemen statistik ekonomi dan moneter dari Bank Indonesia (2004), PDRB terdiri dari PDRB atas dasar harga berlaku dan PDRB atas dasar harga konstan. PDRB atas dasar harga berlaku menggambarkan nilai tambah barang dan jasa yang dihitung menggunakan harga pada tahun berjalan, sedangkan PDRB atas dasar harga konstan menunjukkan nilai tambah barang dan jasa tersebut yang dihitung menggunakan harga yang berlaku pada satu tahun tertentu sebagai tahun dasar.

Pertumbuhan ekonomi merupakan salah satu indikator kinerja yang menggambarkan hasil dari pembangunan yang telah dicapai. Indikator ini penting bagi daerah karena dapat digunakan sebagai bahan evaluasi bagi pemerintah daerah atas keberhasilan pembangunan yang telah dicapai sekaligus sebagai dasar perencanaan dan pengambilan kebijakan dimasa yang akan datang. Arsyad (2000) dalam skripsi Yeni Dharmayanti (2011) menyatakan bahwa pertumbuhan ekonomi daerah diartikan sebagai kenaikan Produk Domestik Regional Bruto (PDRB) tanpa memandang apakah kenaikan itu lebih besar atau lebih kecil dari tingkat pertumbuhan penduduk atau apakah perubahan struktur ekonomi terjadi atau tidak. Hal ini berarti bahwa pertumbuhan ekonomi daerah secara langsung ataupun tidak langsung akan menciptakan lapangan kerja.

Menurut Nainggolan (2009), mengatakan bahwa terdapat adanya pengaruh PDRB dan jumlah pengangguran yang bersifat positif dalam Teori Pertumbuhan Ekonomi. Dikatakan berpengaruh positif sebab pertumbuhan ekonomi tidak dibarengi oleh peningkatan kapasitas produksi, sehingga jumlah pengangguran tetap meningkat seiring pertumbuhan ekonomi yang berlangsung. Hal ini disebabkan pertumbuhan ekonomi yang meningkat tersebut berorientasi pada padat modal, dimana kegiatan produksi untuk memacu output dan menghasilkan pendapatan yang meningkat lebih diutamakan ketimbang pertumbuhan ekonomi yang berorientasi pada padat karya.

2.2.2 Indeks Pembangunan Manusia (IPM)

Indeks pembangunan manusia memberikan suatu ukuran gabungan tiga dimensi tentang pembangunan manusia diantaranya: panjang umur dan menjalani hidup sehat (diukur dari usia harapan hidup), terdidik (diukur dari tingkat kemampuan baca tulis orang dewasa dan tingkat pendaftaran di sekolah dasar, lanjutan dan tinggi), dan memiliki standar hidup yang layak (diukur dari paritas daya beli/PPP, penghasilan) (UNDP, 2004).

IPM digunakan untuk mengklasifikasikan apakah sebuah negara termasuk kategori negara maju, negara berkembang atau Negara terbelakang. Selain itu indeks ini juga menjadi parameter untuk melihat pengaruh kebijakan ekonomi suatu negara terhadap kualitas rakyatnya. Dan tidak hanya digunakan sebagai tolak ukur pengelompokan suatu Negara tetapi juga dapat digunakan sebagai tolak ukur untuk mengukur dan pengelompokan Subnegara (daerah/ bagian) (Cholili, 2014 : 5) .

Indeks Pembangunan Manusia (IPM) merupakan indikator yang menjelaskan bagaimana penduduk suatu wilayah mempunyai kesempatan untuk mengakses hasil dari suatu pembangunan sebagai bagian dari haknya dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. Nilai IPM menunjukkan seberapa jauh wilayah tersebut telah mencapai sasaran yang ditentukan yaitu angka harapan hidup 85 tahun, pendidikan dasar bagi semua lapisan masyarakat, dan tingkat pengeluaran dan konsumsi yang telah mencapai standar hidup layak. Semakin dekat nilai IPM suatu wilayah terhadap angka 100, maka semakin dekat jalan yang harus ditempuh untuk mencapai sasaran itu.

Todaro (2000) mengatakan bahwa pembangunan manusia merupakan tujuan pembangunan itu sendiri. Pembangunan manusia memainkan peranan kunci dalam membentuk kemampuan sebuah negara dalam menyerap teknologi modern dan untuk mengembangkan kapasitasnya agar tercipta pertumbuhan serta pembangunan yang berkelanjutan. Kualitas Sumberdaya Manusia yang dapat dilihat dari nilai Indeks

Pembangunan Manusia dapat menjadi penyebab terjadinya penduduk miskin. Rendahnya Indeks Pembangunan Manusia (IPM) akan berakibat pada rendahnya produktivitas kerja yang berimbas pada rendahnya perolehan pendapatan.

2.2.3 Partisipasi Angkatan Kerja

Adapaun salah satu Indikator pengangguran adalah partisipasi angkatan kerja. Partisipasi angkatan kerja adalah perbandingan antara jumlah angkatan kerja dengan penduduk dalam usia kerja dalam kelompok yang sama. Partisipasi angkatan kerja dapat dinyatakan untuk seluruh penduduk dalam usia kerja dan dapat pula dinyatakan untuk suatu kelompok penduduk tertentu.

Partisipasi angkatan kerja berarti keikutsertaan dalam atau menjadi angkatan kerja. Jadi, tingkat partisipasi angkatan kerja (*labor force participation rate*) menunjuk kepada persentase jumlah penduduk usia kerja yang termasuk dalam angkatan kerja. Sebaliknya partisipasi angkatan kerja berarti keikut sertaan dalam atau mempunyai pekerjaan. Jadi, tingkat partisipasi angkatan kerja menunjuk kepada persentase jumlah angkatan kerja yang mempunyai pekerjaan (*employment rate*).

Penduduk merupakan sumber yang pokok bagi persediaan tenaga kerja. Bagian penduduk yang tidak mempunyai pekerjaan dan juga tidak mencari pekeerjaan dianggap ada di luar angkatan kerja. Sedangkan angkatan kerja terdiri dari mereka yang mempunyai pekerjaan, mereka yang secara aktif mencari pekerjaan dan mereka yang dianggap mencari pekerjaan. Yang terakhir ini dalam survei-survei sering dijumpai sebagai mereka yang sedang tidak mempunyai pekerjaan, tapi tidak mencari pekerjaan karena beranggapan bahwa tidak ada kesempatan kerja yang tersedia baginya. Mereka ini biasanya disebut "*discouraged workers*" atau "penganggur putus asa".

2.3 Analisis Regresi

Analisis regresi merupakan salah satu analisis statistik yang sering digunakan untuk menganalisis hubungan antara dua variabel atau lebih. Analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel terhadap variabel lainnya (Draper dan Smith, 1992). Hubungan yang didapat pada umumnya dinyatakan dalam bentuk persamaan matematika yang menyatakan hubungan antara variabel independen dalam bentuk persamaan sederhana.

$$y = \beta_0 + \beta_1 x + e \tag{2.1}$$

dengan :

y : variabel respon yang akan diteliti

β_0 : konstanta

β_1 : parameter variabel independen

x : variabel prediktor

e : variabel *error* acak

Hubungan atau pengaruh dua atau lebih variabel independen terhadap satu variabel dependen maka model regresi yang digunakan adalah model regresi linear berganda. Persamaan umum regresi berganda dapat ditulis dalam persamaan berikut

(Walpole dan Myers, 1995):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1,2,3, \ldots, n \tag{2.2}$$

dengan:

Y_i : variabel dependen ke-i

β_0 : konstanta

β_k : parameter variabel independen ke-ik

x_{ik} : variabel independen ke-ik

ε_i : variabel *error* acak

Dalam notasi matriks, persamaan (2.2) dapat ditulis sebagai berikut (Montgomery, 2006: 68):

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{2.3}$$

atau dapat ditulis menjadi:

$$\boldsymbol{y_i} = \boldsymbol{x'\beta} + \boldsymbol{\varepsilon}$$

dengan:

y_i : vektor amatan yang berukuran $(n \times 1)$

x' : matriks berukuran $(n \times (k+1))$ yang diketahui

β : vektor parameter yang berukuran $((k+1) \times 1)$

ε : vektor residual yang berukuran $(n \times 1)$

2.4 Metode Kuadrat Terkecil (MKT)

Metode Kuadrat Terkecil merupakan salah satu metode untuk mengestimasi parameter pada regresi linear yaitu $\beta_0, \beta_1, . . \beta_k$. Tujuan MKT adalah meminimumkan jumlah kuadrat dari kesalahan yang disebut dengan jumlah kuadrat galat terhadap garis regresi.

Nilai estimasi $\hat{\beta}$ dapat dicari dengan menggunakan persamaan berikut :

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \tag{2.4}$$

Pada penerapan nilai estimasi $\hat{\beta}$ dapat dilakukan baik pada regresi sederhana mapun regresi berganda. Akan tetapi biasanya matriks lebih sering digunakan untuk regresi linier berganda.

2.5 Uji Asumsi Klasik

Model regresi yang diperoleh dari metode kuadrat terkecil merupakan model regresi yang menghasilkan estimator linear tidak bias dan yang terbaik (*BLUE*). Kondisi ini akan terjadi jika dipenuhi beberapa asumsi yang disebut asumsi klasik (Algifari, 2000). Terdapat beberapa asumsi yang harus dipenuhi terlebih dahulu sebelum menggunakan *Multiple Linear Regression* sebagai alat untuk menganalisis pengaruh variabel-variabel yang diteliti, pengujian asumsi klasik yang digunakan terdiri dari:

2.5.1 Uji Autokorelasi

Uji autokorelasi bertujuan untuk menguji apakah dalam model regresi linier ada korelasi antara kesalahan pengganggu pada periode sebelumnya. Autokorelasi ini muncul karena observasi yang berurutan sepanjang waktu berkaitan satu sama lainnya. Korelasi antar observasi ini diukur berdasarkan deret waktu dalam model regresi atau dengan kata lain *error* dari observasi yang satu dipengaruhi oleh *error* dari observasi yang sebelumnya atau dapat dikatakan bahwa penyimpangan asumsi ini biasanya muncul pada observasi yang menggunakan data *time series.* (Algifari, 2000). Hipotesis yang digunakan sebagai dugaan awal adalah:

H_0: tidak terdapat autokorelasi

H_1: terdapat autokorelasi

Ada tidaknya gejala autokorelasi dapat dideteksi dengan uji *Durbin Watson* yang dirumuskan sebagai berikut:

$$d = \frac{\sum_{i=2}^{N}(e_i - e_{i-1})^2}{\sum_{i=1}^{N} e_i^2} \quad (2.5)$$

Setelah nilai *Durbin Watson* dihitung, pengambilan keputusan dilakukan dengan membandingkan nilai batas d_L dan d_U untuk berbagai jumlah sampel (n) dan jumlah variabel bebas (k) yang ada dalam tabel *Durbin Watson*. Berikut ketentuan pengambilan keputusan uji autokorelasi berdasarkan (Gujarati, 2013):

a. Jika $d < d_L$ atau $d > 4 - d_L$ maka kesimpulannya adalah tolak H_0 artinya terdapat autokorelasi antara *residual*.
b. Jika $d_U < d < 4 - d_U$ maka kesimpulannya adalah gagal tolak H_0 atau tidak terdapat autokorelasi antara *residual*.
c. Jika $d_L \leq d \leq d_u$ atau $4 - d_U \leq d \leq 4 - d_L$ pengujian tidak meyakinkan sehingga tidak dapat disimpulkan ada atau tidaknya autokorelasi antara *residual.*

2.5.2 Uji Heterokedastisitas

Uji Heteroskedastisitas bertujuan menguji apakah dalam model regresi terjadi ketidaksamaan variansi dari residual satu pengamatan ke pengamatan yang lain. Jika variansi dari residual satu pengamatan ke pengamatan lain tetap, maka disebut homokedastisitas dan jika berbeda disebut heteroskedastisitas.

Model regresi yang baik adalah yang homokedastisitas atau tidak terjadi heteroskedastisitas (Ghozali, 2009). Uji Heteroskidastisitas bertujuan untuk mengetahui varians variabel dalam model tidak sama (konstan). Situasi heterokedastisitas akan menyebabkan penaksiran koefisien-koefisien regresi menjadi tidak efisien dan hasil taksiran dapat menjadi kurang atau melebihi dari yang semestinya. Dengan demikian, agar koefisien-koefisien regresi tidak menyesatkan maka situasi heteroskedastisitas tersebut harus dihilangkan dari model regresi. Uji heteroskedastisitas dapat dilakukan dengan menggunakan uji *Glejser* yaitu dengan meregresikan nilai *absolute* residual dengan masing-masing variabel independen. Hipotesis yang digunakan sebagai dugaan awal adalah sebagai berikut:

H_0: tidak terdapat heteroskedastisitas

H_1: terdapat heteroskedastisitas

Keputusan H_0 dilakukan ketika *p-value* > α: 5% yang artinya model dinyatakan bebas masalah heteroskedastisitas.

2.5.3 Uji Multikolinearitas

Uji Multikolinieritas bertujuan untuk mengetahui ada tidaknya hubungan antara variabel independen atau bebas. Pengujian ini dilakukan sebagai syarat digunakannya analisis berganda dimana regresi yang baik adalah regresi yang terbebas dari masalah multikolinearitas. (Ghozali, 2009).

Penentuan ada atau tidaknya multikolinearitas salah satunya dengan melihat nilai *Variance Inflation Factor* dan nilai *tolerance* dengan hipotesis sebagai berikut:

H_0: tidak terdapat multikolinearitas

H_1: terdapat multikolinearitas

Gejala multikolinearitas terjadi ketika $VIF \geq 10$ dan nilai *tolerance* $\leq 0{,}10$ (Ghozali, 2009). Perhitungan nilai *VIF* dapat dilakukan dengan menggunakan persamaan berikut (Montgomery, Peck & Vining, 1982):

$$VIF = \frac{1}{1-R^2} \qquad (2.6)$$

dengan R^2 :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} \qquad (2.7)$$

dengan $i = 1,2,3, \ldots k$ dan R^2 adalah koefisien regresi yang dihasilkan dari variabel independen x_i dengan variabel independen lain x_j $(x_i \neq x_j)$.

2.5.4 Uji Normalitas Residual

Uji normalitas digunakan untuk menguji apakah model regresi variabel bebas dan variabel terikat keduanya mempunyai distribusi normal atau tidak (Ghozali, 2009). Pada regresi linear diasumsikan bahwa tiap e_i didistribusikan secara random dengan $e_i \sim N(0, \sigma^2)$. Jika asumsi ini tidak terpenuhi akan menyebabkan residual yang besar. Salah satu uji yang dapat digunakan adalah uji *Kolmogorov Smirnov*. Uji ini secara sistematis sebagai berikut:

$$D = \max |F_0(X_i) - F_n(X_i)| \quad (2.8)$$

dengan $i = 1,2, \ldots, n$, $F_0(X_i)$ adalah fungsi distribusi frekuensi kumulatif relatif dari distribusi normal dan $F_n(X_i)$ adalah distribusi frekuensi kumulatif pengamatan sebanyak n sampel dengan menggunakan kurva normalitas. Adapun hipotesis dalam pengujian kenormalan adalah sebagai berikut:

H_0 : Data berdistribusi normal

H_1 : Data tidak berdistribusi normal

Apabila nilai $D > D_{tabel}$ atau $P_{value} < \alpha$ maka asumsi normalitas residual tidak terpenuhi.

2.6 Data Outlier

Outlier adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan atau tidak mengikuti pola data secara keseluruhan. Dalam suatu himpunan data biasanya terdapat 10% amatan yang merupakan *outlier* (Rousseeuw dan Leroy 1987). Keberadaan *outlier* akan mengganggu dalam proses analisis data dan harus dihindari dalam banyak hal. Dalam kaitanya dengan analisis regresi. *Outlier* menyebabkan hal-hal berikut (Soemartini, 2007):

1. Residual yang besar dari model yang terbentuk
2. Varians pada data tersebut menjadi besar
3. Taksiran interval memiliki rentang yang lebar

Pada analisis regresi, terdapat 3 tipe *outlier* yang berpengaruh terhadap estimasi kuadrat terkecil yaitu sebagai berikut (Soemartini, 2007):

a. Outlier Vertikal *(Vertical Outlier)*

Merupakan semua pengamatan yang terpencil pada variabel dependen, tetapi tidak terpencil pada variabel independen. Keberadaan *vertical outlier* mempengaruhi terhadap estimasi kuadrat terkecil yang dapat digambarkan pada Gambar 1.

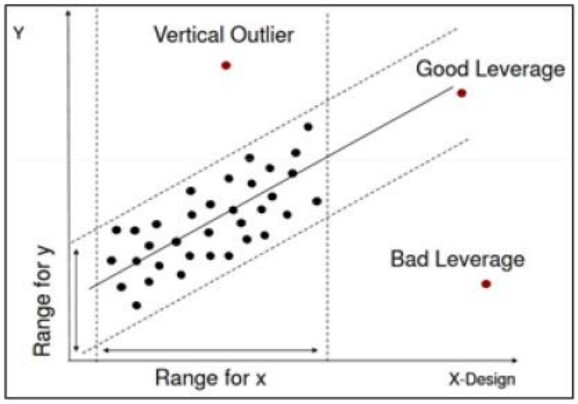
b. Good Leverage Point

Merupakan pengamatan yang terpencil pada variabel independen tetapi terletak dekat dengan garis regresi. Hal ini berarti pengamatan menjauh tetapi dekat dengan garis regresi. Keberadaan *good leverage point* tidak berpengaruh terhadap estimasi kuadrat terkecil, tetapi berpengaruh terhadap inferensi statistik karena dapat meningkatkan estimasi standar *error* seperti Gambar 1.

c. Bad leverage point

Merupakan pengamatan yang terpencil pada variabel independen dan terletak jauh dari garis regresi. Keberadaan *bad leverage point* berpengaruh signifikan terhadap estimasi kuadrat terkecil, baik terhadap *intercept* maupun *slope* dari persamaan regresi seperti Gambar 1.

Apabila dalam pengamatan terdapat data *outlier* maka data dilakukan tranformasi, akan tetapi menghapus *outlier* memberikan informasi yang tidak bias diberikan oleh titik data lainnya, misalnya *outlier* timbul karena kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh (Draper & Smith, 1992).



Gambar 2.1 Tipe Data Outlier
(Sumber: Verardi, 2008)

2.7 Identifikasi Data *Outlier*

Metode identifikasi *outlier* terbagi menjadi dua yaitu metode grafis yang hanya mengandalkan visualisasi dan sangat bergantung kepada cara pandang peneliti terhadap grafik yang terbentuk sehingga tetap perlu dilakukan metode kedua yaitu metode perhitungan statistik. Beberapa metode identifikasi *outlier* dalam sebuah analisis adalah sebagai berikut:

a. *Scatterplot*

Metode ini dilakukan dengan cara memplotkan data dengan observasi ke$-i$ (i= 1,2, ... n). Selain itu, jika sudah didapatkan model regresi maka dapat dilakukan dengan cara memplot residual (e) dengan nilai prediksi Y $(\hat{Y})$. Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan maka hal ini mengindikasikan adanya *outlier*.

b. Metode *Leverage Values*

Metode ini mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter. Hal ini dapat dilihat dari jarak nilai X semua observasi. Nilai *leverage* untuk linear sederhana dapat ditentukan sebagai berikut (Wijaya, 2009):

$$Leverage\ (h_{ii}) = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n-1)S_x^2} \qquad (2.9)$$

dengan:

h_{ii} = *leverage* kasus ke-i

n = banyaknya data

X_i = nilai untuk kasus ke-i

S_x^2 = kuadrat n kasus terdiri dari simpangan Xi terhadap *mean*

$\bar{X}$ = *mean* dari X

Jika suatu kasus terdiri dari beberapa variabel independen maka perhitungan nilai *leverage* dapat dilakukan dengan menggunakan persamaan matriks berikut:

$$H = X(X'X)^{-1}X' \qquad (2.10)$$

dengan H adalah *hat* matriks, elemen ke-i dari diagonal dari *hat* matriks merupakan *leverage* dan X merupakan matriks X. Pendekatan *outlier* berdasarkan pada nilai *cutoff* dan apabila nilai h_{ii} melebihi nilai *cutoff* dideteksi sebagai *outlier*. Adapun nilai *cutoff* yang telah ditentukan adalah $\frac{2p}{n}$, dengan n merupakan banyaknya data, dan p merupakan banyaknya parameter pada persamaan regresi yang terbentuk termasuk *intercept* (Kurtner, 2004).

e. Metode *DfFITS (Difference fitted Value FITS)*

Metode ini menampilkan nilai perubahan dalam harga yang diprediksi ketika kasus ke-i dihapuskan dalam penelitian data penelitian yang sudah distandarkan (Soemartini, 2007). Perhitungan *DfFITS* adalah sebagai berikut:

$$DfFITS = t_i \left(\frac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}} \tag{2.11}$$

dimana t_i adalah *studentized deleted residual* untuk kasus ke-i dan h_{ii} adalah nilai *leverage* untuk kasus ke-i. Data dikatakan *outlier* apabila nilai $|DfFITS| > 2\sqrt{\frac{p}{n}}$ dengan p adalah banyaknya parameter dan n adalah banyaknya data observasi.

2.8 Regresi *Robust*

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terkontaminasi oleh *outlier* dan memberikan hasil yang lebih fleksibel. Regresi *robust* tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang kecil untuk data pencilan (Soemartini, 2007). Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil terhadap adanya *outlier* (Chen, 2002).

Terdapat dua hal penting yang sangat diperlukan dalam estimasi regresi *robust* yaitu resistensi dan efisiensi. Suatu estimasi dikatakan resisten terhadap *outlier* jika sebagian kecil dari data tidak dapat memberikan efek yang terlalu besar terhadap estimator sedangkan estimasi memiliki efisiensi yang cukup baik pada berbagai sebaran jika ragamnya mendekati ragam minimum untuk setiap sebaran (Mortgomery & Peck, 1982). Pada regresi *robust* terdapat beberapa metode estimasi yang dapat digunakan, yakni (1) estimasi *LMS*, (2) estimasi *LTS*, (3) estimasi *MM*, (4) estimasi *S*, (5) estimasi *M* (Chen, 2002).

2.8.1 Estimasi LTS (Least Trimmed Square)

Estimasi *LTS* merupakan metode penduga regresi *robust* yang menggunaan konsep pengepasan metode kuadrat terkecil untuk meminimumkan jumlah kuadrat sisaan (Akbar & Maftukhah, 2007). Penduga *LTS* (β) dinyatakan dalam bentuk rumus sebagai berikut. (Rousseeuw & Leroy, 1987):

$$\hat{\beta}_{LTS} = \arg min \sum_{i}^{h} e_i^2 \tag{2.12}$$

dengan:

$h = \frac{n+p+2}{2}$; $e_i = (y_i - x_i'\beta)$

$e_1^2 \leq e_2^2 \leq \cdots \leq e_n^2$ = sisaan kuadrat yang diurutkan dari terkecil ke terbesar

n = banyaknya sampel

p = banyaknya parameter

Jumlah h menunjukan sejumlah subset data dengan kuadrat fungsi obyektif terkecil. Nilai h pada persamaan akan membangun *breakdown point* yang besar sebanding dengan 50%. Kuadrat sisa berasal dari persamaan estimasi regresi linear menggunakan konsep metode kuadrat terkecil dengan banyaknya sisaan kuadrat (e^2) yang akan diolah adalah sebanyak h residual. Estimator berdasarkan pada estimasi S_{LTS} disebut juga sebagai *Final Weight Scale Estimator (FWLS)*. Secara matematis fungsi pembobotnya jika nilai r =3 sebagai berikut:

$$w_i = \begin{cases} 0 & , \frac{|e_i|}{S_{LTS}} \\ 1 & , lainnya \end{cases} \tag{2.13}$$

dengan :

$$S_{LTS} = d_{h,n} \sqrt{\frac{1}{h} \sum_{i}^{h} e_{(i)}^2} \tag{2.14}$$

$$d_{h,n} = \frac{1}{\sqrt{1\frac{2n}{hC_{h,n}}f(z)\left(\frac{1}{C_{h,n}}\right)}} \quad (2.15)$$

$$C_{h,n} = \frac{1}{f(z)^{-1}\left(\frac{h+n}{2n}\right)} \quad (2.16)$$

dengan:

n : banyaknya observasi

$f(z)$: fungsi *density* normal standar

2.8.2 Estimasi S (Scale)

Pada situasi ketika data terkontaminasi *outlier* pada variabel X . Estimasi M tidak dapat bekerja dengan baik. Estimasi M tidak dapat mengidentifikasi *bad observation* yang berarti tidak dapat membedakan *good leverage point* dan *bad leverage point* dan untuk mengatasi hal tersebut, estimasi *high breakdown* sangat diperlukan. (Chen, 2002). Salah satu estimasi yang mempunyai nilai *high breakdown* adalah estimasi S.

Estimasi S akan meminimumkan jumlah kuadrat *error* pada persamaan umum regresi linear. Estimasi S didefinisikan sebagai berikut :

$$\hat{\beta}_s = arg_{\beta}\min S_s(e_1 e_2, \dots, e_n) \quad (2.17)$$

Dimana S_s adalah estimator skala *robust* yang memenuhi $\frac{1}{n}\sum_{i=1}^{n} \rho \frac{e_i}{\hat{\sigma}_s} = b$ dengan b konstan yang didefenisikan $b = E[z, (\rho)]$, $f(z)$ adalah distribusi normal standar. Estimator S mempunyai nilai *breakdown* tinggi yaitu 50%. Nilai *breakdown* dari estimator S ditulis $\frac{b}{\max(\rho(e))} = 0{,}5$ dengan S_s adalah nilai estimator skala *robust* yang minimum dan memenuhi:

$$\min \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{j=0}^{k} x_i \beta_j}{S_s} \right) \tag{2.18}$$

dengan :

$$S_s = \sqrt{\frac{1}{nK} \sum_{i=1}^{n} w_i e_i^2} \tag{2.19}$$

$K = 0.199, w = w_s(u_i) = \frac{\rho(u_i)}{u_i^2}$ dan estimasi awal yang digunakan adalah

$$S_{awal} = \frac{median|e_i - median(e_i)|}{0.6745} \tag{2.20}$$

Estimator $\hat{\beta}$ pada metode regresi *robust* estimasi S diperoleh dengan cara melakukan iterasi hingga diperoleh hasil yang konvergen. Proses ini dikenal sebagai MKT terboboti secara iterasi yang selanjutnya disebut sebagai *Iteratively Reweighted Least Squared (IRLS)* (Fox & Weisberg, 2010).

2.8.3 Estimasi M (Maximum Likelihood Type)

Metode penaksiran M merupakan metode penaksiran dalam regresi *robust* untuk mengestimasi parameter yang disebabkan adanya *outlier*. Penaksiran M meminimumkan fungsi ρ (fungsi obyektif) dari residualnya. Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust.* Fungsi pembobot yang digunakan adalah fungsi pembobot *Huber* dan fungsi pembobot *Tukey* (Montgomery & Peck, 1982).

$$\min_{\beta} \rho\,(e_i) = \min_{\beta} \sum_{i=1}^{n} \rho(y_i - \sum_{j=0}^{k} x_{ij} \beta_j) \tag{2.21}$$

Penaksiran parameter menggunakan metode penaksiran M disebut *Iteratively Raweighted Least Square (IRLS).* Mengestimasi parameter regresi pada regresi *robust* menggunakan esimasi M dilakukan iterasi hingga diperoleh nilai estimasi parameter yang konvergen.

2.8.4 Breakdown Point

Breakdown point adalah *persentase* terkecil dari banyaknya data yang terkontaminasi atau banyaknya *outlier* yang menyebabkan nilai dari taksiran menjadi besar. (Rousseeuw & Leroy, 1987). *Breakdown point* digunakan untuk menjelaskan ukuran ke-*robust*-an dari teknik *robust*. Kemungkinan tertinggi *breakdown point* untuk sebuah estimator adalah 50%. Jika *breakdown poin*t lebih dari 50% berarti estimasi model regresi tidak dapat menggambarkan informasi dari kebanyakan data.

2.8.5 Koefisien Determinasi

Koefisien determinasi atau *Adj-Square* merupakan salah satu ukuran yang sederhana dan sering digunakan untuk menguji kualitas suatu persamaan garis regresi (Gujarati, 1978). Nilai koefisien determinasi digunakan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen. Semakin besar nilai *Adj-Square*, maka semakin besar variasi variabel dependen yang dijelaskan oleh variasi variabel independen. Sebaliknya, semain kecil nilai *Adj-Square* maka semakin kecil variansi variabel dependen yang dijelaskan oleh variabel independen, namun salah satu kelemahan R adalah besarnya dipengaruhi oleh banyaknya peubah bebas dalam model. Sehingga sulit menyatakan berapa R^2 yang optimum. Akan tetapi nilai model yang ingin dibandingkan mempunyai jumlah variabel independen yang sama maka R^2 mudah digunakan, yaitu dengan memilih R^2 terbesar (Sembiring, 2003).

Suatu cara mengatasi kelemahan R^2 tersebut di atas ialah dengan menggunakan apa yang disebut $\bar{R}^2$. berikut perhitungan $\bar{R}^2$:

$$\bar{R}^2 = 1 - \left((1 - R^2) \text{ x } \frac{n-1}{n-k}\right) \quad (2.22)$$

dengan n adalah banyaknya data observasi dan k adalah banyaknya variabel dan R^2 dapat dilihat pada persamaan 2.7.

2.8.6 Mean Squared Error (MSE)

Terdapat beberapa cara untuk melihat kebaikan pendugaan parameter regresi yaitu dengan melihat niai *MSE*. Model persamaan yang baik adalah model regresi dengan nilai *MSE* kecil. Semakin kecil nilai *MSE* yang dihasilkan maka semakin baik pendugaan parameter yang dihasilkan tersebut. Nilai *MSE* diperoleh dari nilai jumlah kuadrat galat dibagi dengan *db* jumlah kuadrat sisaan. Berikut perhitungan nilai *MSE* (Sembiring, 2003):

$$MSE = \frac{JKG}{n-k} \quad (2.23)$$

dengan:

$$JKG = \sum_{i=n}^{n} \left(Y_i - \hat{Y}_i\right)^2 \quad (2.24)$$

dengan JKG adalah jumlah kuadrat galat, n adalah jumlah sampel dan p adalah jumlah parameter yang diestimasi.

2.9 Pengujian Signifikansi Parameter

Pengujian signifikansi parameter dibagi menjadi dua yaitu uji *overall* (serentak) dan uji *parsial* (individu), (Mortgomery, Leroy dan Vining, 1982):

2.9.1 Uji Overall

Uji *overall* atau uji serentak atau uji F merupakan pengujian untuk mengetahui ada atau tidaknya pengaruh secara bersama-sama variabel prediktor terhadap variabel respon. Hipotesis dalam pengujian ini adalah sebagai berikut:

H_0: $\beta_1 = \beta_2 = \cdots \beta_k = 0$

H_1: Minimal ada satu $\beta_j \neq 0$ untuk $j = 0, 1, \ldots k$

Dasar pengambilan keputusannya yaitu apabila $F_{hitung} > F_{tabel}$ maka H_0 ditolak. Artinya minimal ada satu B_j yang tidak sama dengan nol. Selain menggunakan dasar keputusan tersebut dapat digunakan dasar pengambilan keputusan lain yaitu jika $P - value < a$ maka H_0 ditolak.

2.9.2 Uji Parsial

Uji *parsial* atau uji t adalah pengujian masing-masing variabel x_i terhadap model. Tujuan dari uji *parsial* ini adalah untuk mengetahui adanya pengaruh antara variabel prediktor ke-j dengan $j = 0, 1, \ldots k$ dengan variabel respon. Hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut:

H_0: $\beta_j = 0$

H_1: $\beta_j \neq 0$ untuk $j = 0, 1, \ldots k$

Statistik uji yang digunakan adalah sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\beta_j)} \tag{2.25}$$

dengan $SE(\beta_j)$ adalah nilai standar *error* dari β_j.

Pengambilan keputusan statistik uji tersebut adalah apabila $|t_{hitung}| > t_{(1-(\frac{\alpha}{2}), n-k-1)}$ dengan k adalah jumlah parameter, maka tolak H_0 yang artinya terdapat pengaruh variabel *independent* terhadap model.

2.9.3 Fungsi Objektif dan Fungsi Pembobot

Fungsi objektif merupakan representasi pembobot dari residual atau $\rho(u)$, fungsi ini digunakan untuk mencari fungsi pembobot pada regresi *robust* sedangkan fungsi pembobot didapatkan dengan menggunakan fungsi objektif. Secara ringkas, fungsi objektif dan pembobot dari estimator MKT dan pembobot Huber dapat dilihat pada Tabel 2.1 sebagai berikut:

Tabel 2.1 Fungsi Objektif dan Fungsi Pembobot

Metode	**Fungsi Objektif**	**Fungsi Pembobot**	**Interval**
MKT	$\rho(u_i) = \frac{1}{2}u_i^2$	$w(u_i) = 1$	$\lvert u_i \rvert < \infty$
Huber	$\rho(u_i) = \begin{cases} \frac{1}{2}u_i^2 \\ r\lvert u_i \rvert - \frac{r^2}{2} \end{cases}$	$w(u_i) = \begin{cases} 1 \\ \frac{r}{\lvert u_i \rvert} \end{cases}$	$\lvert u_i \rvert \leq r$ $\lvert u_i \rvert > r$

Sumber: Mortgomery, Leroy dan Vining, 1982