BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan salah satu metode statistika yang mempelajari persamaan secara matematis hubungan antara satu peubah respon dengan satu atau lebih peubah penjelas. Draper dan Smith (2014) mendefinisikan hubungan antara peubah respon dan peubah penjelas dalam model regresi linear. Secara umum dituliskan dalam persamaan sebagai berikut :

$$\boldsymbol{Y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,1} + \beta_2 \boldsymbol{X}_{i,2} + \ldots + \beta_{p-1} \boldsymbol{X}_{i,p-1} + \boldsymbol{\varepsilon}_i \qquad (1)$$

Dimana Y_i merupakan peubah respon untuk pengamatan ke-i. $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ adalah parameter peubah penjelas. Peubah penjelas di tuliskan dalam $\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, \ldots, \boldsymbol{X}_{i,p-1}$ dan $\boldsymbol{\varepsilon}_i$ adalah sisa untuk pengamatan ke-*i* yang diasumsikan berdistribusi normal yang saling bebas dan identik dengan rata-rata 0 (nol) dan varians σ^2 . Secara ringkas persamaan di atas dapat ditulis menjadi persamaan (2):

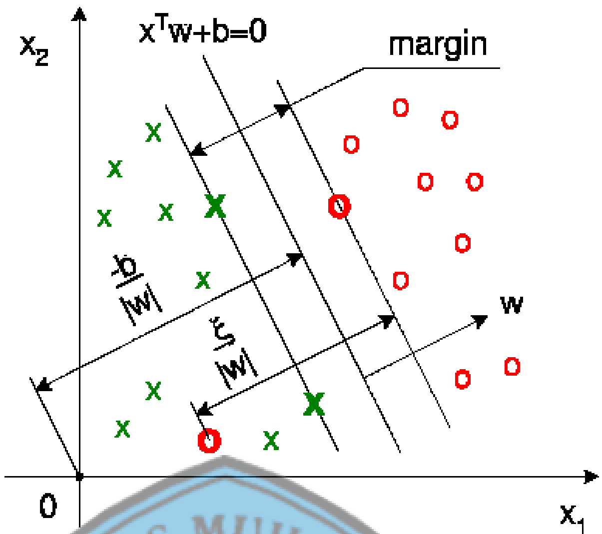
$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2)$$

Dengan **Y** dituliskan sebagai vektor peubah respon berukuran *n* x 1, **X** merupakan matriks peubah penjelas berukuran *n* x (*p* – 1), $\boldsymbol{\beta}$ adalah vektor parameter berukuran *p* x 1, dan ε merupakan vektor sisaan berukuran *n* x 1.

2.2 *Support Vector Machine*

Support Vector Machine adalah suatu algoritma yang tergolong dalam mesin pembelajaran (*machine learning*). Dasar dari algoritma ini dikembangkan oleh Vapnik dan Cortes (1995). Metode SVM merupakan metode dalam klasifikasi yang digunakan untuk menyelesaikan masalah linear dan mulai dikembangkan untuk memecahkan masalah nonlinear. Metode SVM mempunyai akurasi yang tinggi, dengan kemampuannya dalam membuat model nonlinear yang kompleks. Konsep yang dipakai dalam SVM adalah dengan menentukan *hyperlane* atau garis yang memisahkan antara satu kelompok dengan kelompok lainnya. SVM menemukan *hyperplane* menggunakan *support vector* dan *margin* (Pudjo *et al*., 2013).

Metode SVM mencari *hyperplane* dengan *margin* yang paling besar yang disebut dengan *maximum marginal hyperplane* (MMH). *Margin* merupakan jarak antara hyperplane dengan data terdekat dalam masing-masing kelas. Nilai *margin* menentukan jarak antar kelas. *Support vector* adalah data terdekat dengan *hyperplane* atau berada tepat diatas *hyperplane* pada masing-masing kelas. Hal itu berarti data tersebut berada dekat dengan MMH. SVM digunakan untuk menyelesaikan masalah linear sering disebut *Linear Separable Data* yaitu data yang dapat dipisahkan secara linear. Diasumsikan terdapat data latih $\{x_i, y_i\}$, x_i merupakan atribut untuk data latih $\{x_1, ..., x_n\}$ dan $y_{i \in}\{-1,1\}$ adalah label kelas dari data latih x_i. Pada gambar 2.2 dibawah ini dapat dilihat bidang pemisah yang bisa memisahkan data. Dari gambar juga terlihat data yang berada dekat dan diatas bidang pemisah atau *hyperplane. Hyperplane* dinyatakan sebagai :



Gambar 2.1 *hyperlane* terbaik

$$wx + b = 0 \qquad (3)$$

Dalam gambar 2.1 diatas terdapat dua kelas yang dipisahkan oleh dua bidang pembatas secara sejajar. Kelas pertama yaitu +1 dan kelas kedua yaitu -1. Data pada kelas pertama disimbolkan dengan tanda silang berwarna hijau, sedangkan data pada kelas kedua disimbolkan dengan lingkaran berwarna merah. Persamaan yang diperoleh :

$$wx_i + b \geq +1 \text{ for } y_i = +1 \qquad (4)$$

$$wx_i + b \geq -1 \text{ for } y_i = -1 \qquad (5)$$

Dimana w merupakan *hyperplane* normal. $\frac{-b}{||w||}$ merupakan jarak lurus dari *hyperplane* ke titik pangkal yaitu x=0 dan y=0. Dengan menggabungkan persamaan 4 dan 5 diatas, bisa didapatkan :

$$y_i\left(wx_i+b\right)-1\geq 0 \quad (6)$$

Dalam kasus klasifikasi linier SVM ketika terdapat data yang tidak dapat dikelompokkan dengan benar (*nonseparable case*), rumusan SVM ditambah dengan adanya variabel slack (Nuha, 2012). Formulasi dari permasalahan sebelumnya kemudian diubah menjadi berikut :

$$\min_{w,b,\xi}\frac{1}{2}\|w\|^2+c\sum_{i=1}^{l}\xi_i \quad (7)$$

Dengan kendala $y_i(\boldsymbol{w}.\boldsymbol{x_i}+b)+\xi_i\geq 1, \xi_i\geq 0, i=1,2,\dots.l$ dimana ξ_i adalah variabel slack yang digunakan untuk memberikan penalti pada data yang tidak memenuhi persamaan *hyperplane* $y_i(\boldsymbol{w}.\boldsymbol{x_i}+b)\geq 1$. Untuk meminimalkan nilai variabel slack, pada rumusan diberikan penalty dengan menambah nilai *cost* (C). Parameter C berguna untuk mengontrol pertukaran antara *margin dan error* klasifikasi (Prasetyo, 2012). Menurut Tan *et al.* (2006) dalam Nuha, dkk (2012), untuk kasus nonseparable case, kendala $\propto_i\geq 0$ diubah menjadi $0 \leq\propto_i\geq C_I$.

2.3 *Support Vector Regression*

Support Vector Regression (SVR) merupakan metode pengembangan SVM untuk mengatasi kasus regresi yang menghasilkan output data berupa bilangan riil atau kontinyu (Ju *et al*., 2013). Metode SVM memetakan vektor input dengan cara non linier ke dalam ruang fitur berdimensi tinggi yang kemudian diterapkan pada SVR (Vapnik dan Cortes, 1995). Konsep SVR didasarkan pada *risk minimization*, yaitu untuk mengestimasi suatu fungsi dengan cara meminimalkan batasatas dari *generalization error*, sehingga SVR mampu mengatasi *overfitting* (Yasin *et al*., 2014). Fungsi regresi dari metode SVR adalah :

$$f(x) = w^T \varphi(x) + b \tag{8}$$

Dengan w merupakan vektor pembobot, $\varphi(x)$ merupakan fungsi yang memetakan x dalam suatu dimensi dan b merupakan *bias*.

Pada regresi terdapat residual misalkan residual tersebut (r) didefinisikan dengan mengurangkan output skalar y terhadap estimasi $f(x)$ yaitu $r = y - f(x)$ dengan :

$$E(r) = \begin{cases} 0 & \text{untuk } |r| \le \varepsilon \\ |r| - \varepsilon & \text{untuk yang lain} \end{cases} \tag{9}$$

Dengan ε adalah nilai positif yang kecil. Misalkan pada regresi linier ditentukan residual dari output y dan estimasi $f(x)$ dengan :

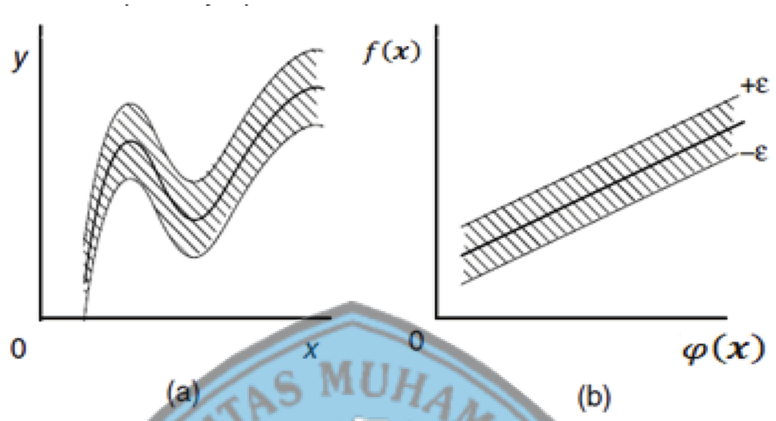
$$r = D(\chi, y) = y - f(x) \tag{10}$$

Berdasarkan persamaan (10) estimasi yang baik akan diperoleh ketika semua nilai dari absolute residual berada pada selang ε.

$$-\varepsilon \le D(\chi, y) \le +\varepsilon \tag{11}$$

Dapat ditulis dengan :

$$|D(\chi, y)| \leq \varepsilon \tag{12}$$



Gambar 2.2 *Insensitive zone* (a) *Original Input Space* dan (b) *Feature Space*

Gambar 2.2 (a) mengilustrasikan *original input* dan *output space*, jika seluruh data training berada diantara radius ε maka akan menghasilkan estimasi yang sesuai. Zona di dalam selang ε disebut *ε-insensitive zone*. Gambar 2.2 (b) merupakan *input-output feature space* dengan dimensi baru sehingga gambar menjadi lebih linier diasumsikan seluruh data training memenuhi pertidaksamaan (12). $D(\boldsymbol{x}, y) = \pm\varepsilon$ merupakan jarak terjauh *support vector* dari *hyperplane*, kemudian disebut margin. Memaksimalkan margin akan meningkatkan probabilitas data ke dalam radius $\pm\varepsilon$, jarak dari *hyperplane* $D(\boldsymbol{x}, y) = 0$ ke data $(\boldsymbol{x}, y)$ adalah $|D(\boldsymbol{x}, y)|/\|\boldsymbol{w}^*\|$, dimana :

$$w^* = (1 - w^T)^T \tag{13}$$

Diasumsikan bahwa jarak maksimum data terhadap hyperplane adalah δ, maka estimasi yang ideal akan terpenuhi dengan :

$$\frac{|D(\chi, y)|}{\|w^*\|} \leq \delta \tag{14}$$

$$|D(\chi, y)| \leq \delta \|w^*\| \tag{15}$$

Pada pertidaksamaan (12) dan (15) data yang terjauh dari hyperplane dipenuhi dengan $|D(\boldsymbol{x}, y)| = \varepsilon$ maka diperoleh :

$$\delta \|w^*\| = \varepsilon \tag{16}$$

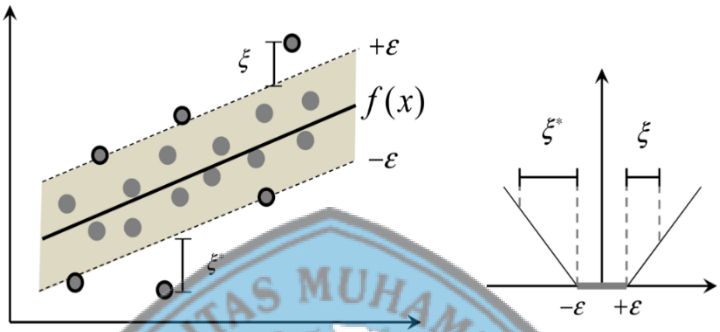
Oleh karena itu untuk memaksimalkan *margin* δ, diperlukan $\|\boldsymbol{w}^*\|$ yang minimum. Sedangkan $\|\boldsymbol{w}^*\|^2 = \|\boldsymbol{w}^*\|^2 + 1$ maka meminimalkan $\|\boldsymbol{w}\|$ juga membuat margin minimal. Optimasi penyelesaian masalah dengan bentuk *Quadratic Programming:*

$$\min \frac{1}{2} \|w\|^2 \tag{17}$$

Factor $\|\boldsymbol{w}\|^2$ dinamakan regulasi. Meminimalkan $\|\boldsymbol{w}\|^2$ akan membuat suatu fungsi setipis mungkin, sehingga bias mengontrol kapasitas fungsi.

Pada persamaan (22) diasumsikan bahwa semua titik ada dalam rentang $f(x) \pm \varepsilon$, dalam hal ketidaklayakan, dimana ada beberapa titik yang mungkin keluar dari rentang $f(x) \pm \varepsilon$ maka ditambahkan variable slack ξ dan ξ^* untuk mengatasi masalah pembatasan yang tidak layak dalam problem optimasi (Santosa,2007).

(a) (b)



Gambar 2.3 (a) SVR *output* dan (b) ε-*insensitive Loss Function*

Gambar 2.3 menjelaskan bahwa semua titik diluar margin akan dikenai pinalti. Selanjutnya problem optimasi di atas bias diformulasikan sebagai berikut :

$$\min \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right) \tag{18}$$

Loss function adalah fungsi yang menunjukkan hubungan antara error dengan bgaiman error ini dikenai pinalti. Perbedaan *loss function* akan menghasilkan SVR yang berbeda (Santosa, 2007). *Loss function* yang paling sederhana adalah ε-*insensitive Loss Function* sebagai sebuah pendekatan Huber's *loss function* yang memungkinkan serangkaian *support vector* akan diperoleh (Gunn, 1998). Formulasi ε-*insensitive Loss Function* adalah :

$$L\varepsilon\left(y_i, f\left(x_i\right)\right) = \begin{cases} 0; \left|y_i - f\left(x_i\right) \leq \varepsilon\right| \\ \left|y_i - f\left(x_i\right) - \varepsilon; lainya\right| \end{cases} \tag{19}$$

Dengan L_ε merupakan $\varepsilon -$ *insensitive loss function*, c dan ε merupakan parameter. Konsep dari kuadratik *loss function* adalah meminimumkan nilai sebagai berikut:

$$R\left(w, \delta, \delta^*\right) = \frac{1}{2}\|w\|^2 + c\left[\sum_{i=1}^{n}\left(\delta_i + \delta_i^*\right)\right] \tag{20}$$

Dengan batasan :

$$w\varphi\left(x_i\right) + b - y_i \leq \varepsilon + \delta_i^*;$$
$$y_i - w\varphi\left(x_i\right) - b \leq \varepsilon + \delta_i \text{ dan } \delta_i, \delta_i^* \geq 0$$

Dengan menggunakan penyelesaian *lagrange* dalam bentuk:

$$L\left(w, b, \delta_i, \delta_i^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*\right) = \frac{1}{2}\|w\|^2 + c\left[\sum_{i=1}^{n}\delta_i + \delta_i^*\right] - \sum_{i=1}^{n}\alpha_i\left[w\varphi\left(x_i\right) + b - y_i + \varepsilon + \delta_i\right] - \sum_{i=1}^{n}\alpha_i^*\left[y_i - w\varphi\left(x_i\right) - b_i + \varepsilon + \delta_i^*\right] - \sum_{i=1}^{n}\left(\beta_i\delta_i + \beta_i^*\delta_i^*\right) \tag{21}$$

Dengan menggunakan pendekatan Karush-Kuhn-Tuck didapatkan sebagai berikut:

$$Q\left(\alpha, \alpha^*\right) = -\frac{1}{2}\sum_{ij=1}^{l}\left(\alpha_i, \alpha_l^*\right)\left(\alpha_j, \alpha_j^*\right)K\left(\mathrm{X}_i - \mathrm{X}_j\right) - \varepsilon\sum_{i=1}^{l}\left(\alpha_i, \alpha_l^*\right) + \sum_{i=1}^{l}y_i\left(\alpha_i, \alpha_l^*\right) \tag{22}$$

Dengan batasan $\sum_{i=1}^{i}(\alpha_i - \dot{\alpha}_\iota) = 0; 0 \leq \alpha_i \leq C; 0 \leq \dot{\alpha}_\iota \leq C$ dimana $K(x_i, x_j)$ merupakan fungsi kernel.

2.4 Fungsi Kernel

SVM nonlinear menggunakan pendekatan kernel untuk data yang kelasnya tidak terdistribusi secara linear. Kernel merupakan suatu fungsi yang memetakan fitur data dari dimensi awal (rendah) ke dalam fitur baru dengan dimensi yang lebih tinggi (Prasetyo, 2012). Macam-macam fungsi kernel antara lain :

a. Linear

$$K(\chi, y) = \chi . y \tag{23}$$

b. Quadratic

$$K(\chi, y) = (\chi . y)^2 \tag{24}$$

c. Polinomial

$$K(\chi, y) = (\chi . y + c)^d \tag{25}$$

d. Raial Basic Function (RBF)

$$\exp\left[\frac{-1}{2\sigma^2}\|x - x_i\|^2\right] \tag{26}$$

e. Sigmoid

$$K(\chi, y) = \tan(\sigma(x.y) + c) \tag{27}$$

Fungsi kernel yang dipilih harus tepat karena sangat penting untuk menentukan fitur baru dimana *hyperplane* akan dicari.

2.5 Algoritma *Grid Search*

Algoritma *grid search* adalah algoritma yang membagi jangkauan parameter yang akan dioptimalkan kedalam grid dan melintasi semua titik untuk mendapatkan parameter yang optimal (Yasin *et al*., 2014). Penggunaan algoritma grid search harus dipandu oleh beberapa metrik kinerja, yang diukur dengan cross-validation pada data training. Sehingga disarankan untuk mencoba beberapa variasi pasangan parameter pada hyperplane SVR (Hsu et al., 2004). Pasangan parameter hasil dari uji *cross validation* dengan akurasi terbaik merupakan parameter yang optimal. Parameter tersebut akan digunakan untuk membentuk mmodel SVR yang akan digunakan untuk memprediksi data testing untuk serta mendapatkan generalisasi tingkat akurasi model.

Cross-validation adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data training dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitungbagian demi bagian, selanjutnya hitung rata-rata *error rate* untuk mendapatkan *error rate* keseluruhan . Dalam *cross-validation* dikenal validasi *leave-one-out* (LOO). Dalam LOO, data dibagi ke dalam 2 subset, subset 1 berisi N-1 data untuk *training* dan satu sisanya untuk *testing* (Leidiyana, 2013).

$$CV = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_{\neq 1} \right)$$

Dengan : $\hat{Y}_{\neq 1}$= Nilai penaksir Y_i dimana pengamatan ke-i dihilangkan dari proses penaksiran

Y_i = Nilai actual *y* pada pengamatan ke-i

Sehingga algoritma yang terbentuk sebagai berikut :

$x = [x_{1,} x_2, \dots, x_n]$: sampel *training*

Grid : *grid search*

LOO : *Leave-one-out*

CV : Jenis penentuan *error* dengan Cross Validation

C : *cost*

ε : *epsilon*

γ : *gamma*

1. Tentukan nilai-nilai parameter kernel (C, γ dan ε) dengan ε bilangan positif kecil
2. Pasangkan nilai masing-masing parameter
3. LOO membagi x menjadi 2 bagian, bagian 1 berisi n-1 untuk *training* dan satu sisanya untuk *testing*
4. Hitung nilai CV untuk memprediksi error dari masing masing pasangan yang terbentuk
5. Hitung rata-rata error dari masing-masing pasangan parameter kernel
6. Grid akan menentukan parameter yang optimal
7. Selesaikan dengan memilih error terkecil dari CV

2.6 Pengukuran Kinerja Prediksi

Dalam regresi ada beberapa ukuran error yang sering dipakai untuk menilai suatu perfomansi fungsi prediksi. Ukuran error yang digunakan dalam penelitian ini adalah nilai MAPE (Mean Absolute Percentage Error). MAPE digunakan untuk menyatakan tingkat kesalahan dari model. Formula dari MAPE dapat dinyatakan sebagai berikut:

$$MAPE = \frac{\sum_{i=1}^{m} APE}{m} \tag{28}$$

Dimana $APE = \frac{\sum_{i=1}^{m} |y_i - \hat{y_i}|}{y_i} \times 100$

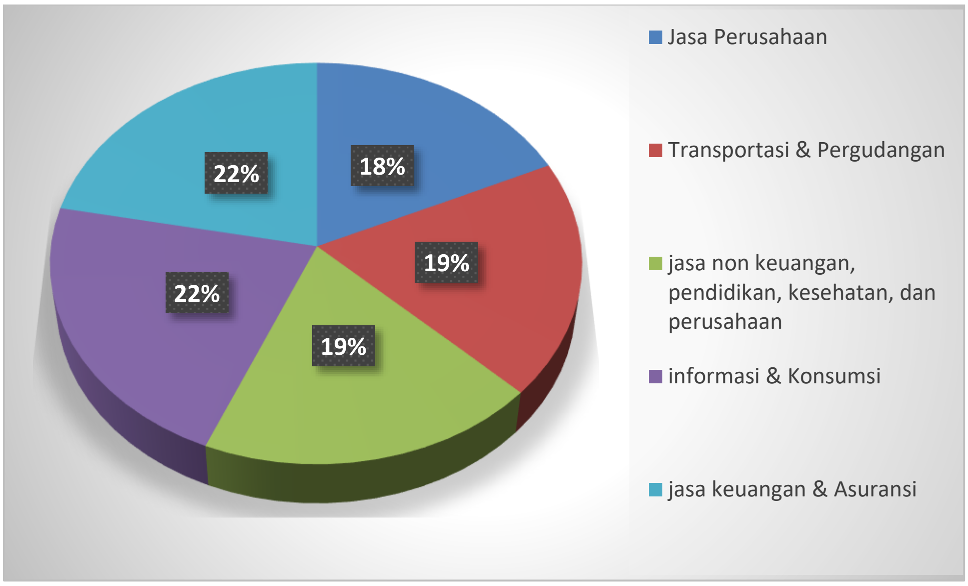
Sedangkan untuk mengukur tingkat akurasi model prediksi menggunakan koefisien determinan R^2.

$$R^2 = 1 - \frac{JKE}{JKT} \tag{29}$$

2.7 Perusahaan Jasa

Sebuah perusahaan tidak hanya menghasilkan produk barang. Karena perkembangan jaman, ilmu pengetahuan serta kebutuhan masyarakat, perusahaan juga menawarkan produk jasa. Perusahaan jasa adalah perusahaan yang kegiatan utamanya memproduksi produk yang tidak berwujud dengan tujuan untuk mencari keuntungan (Alam, 2006). Operasional utama perusahaan jasa adalah memberikan pelayanan, kemudahan serta kenyamanan bagi konsumen guna memperlancar kegiatan produksi maupun konsumsi masyarakat selaku konsumen.

Sektor Jasa atau *services* merupakan salah satu sektor prioritas dalam perekonomian Indonesia, di mana setiap tahunnya kontribusi sektor jasa terhadap PDB Nasional selalu mengalami peningkatan. Mari Elka Pangestu (Board of Advisors Indonesia Services Dialogue (ISD), mengatakan dalam 10 tahun terakhir, kontribusi sektor jasa terhadap PDB terus naik, pada tahun 2000 kontribusi sektor jasa mencapai 45% kemudian meningkat menjadi 60% pada tahun 2015. Dan diperkirakan akan semakin naik untuk tahun berikutnya.



Gambar 2.1 Lima sektor dengan pertumbuhan tertinggi di Indonesia

Pada Gambar 2.1 perusahaan jasa keuangan dan asuransi memiliki pertumbuhan tertinggi di Indonesia, diikuti dengan sector jasa yang lainya (BPS Indonesia). Data tersebut membuktikan bahwa tahun 2016 sektor jasa mengalami pertumbuhan yang tinggi di Indonesia. Pertumbuhan yang tinggi ini dikarenakan kebutuhan akan perusahaan jasa yang semakin meningkat dengan pertumbuhan yang tinggi ini diharapkan akan menjadi penyumbang pertumbuhan ekonomi yang besar pula.

2.8 Perubahan Laba

Menurut Ariyanti (2010) perubahan laba adalah perbedaan antara pendapatan (revenue) yang direalisasi yang timbul dari transaksi pada periode tertentu dengan biaya-biaya yang dikeluarkan pada periode tersebut. Perubahan laba digunakan sebagai indikator kinerja keuangan suatu perusahaan mengalami peningkatan atau penurunan sehingga, perubahan laba akan berpengaruh pada keputusan kebijakan. Informasi perubahan laba juga dapat digunakan untuk memprediksi pertumbuhan laba dimasa mendatang (Ediningsih, 2004). Untuk mengetahui perubahan laba yang terjadi pada perusahaan akan digunakan rumus sebagai berikut :

$$\Delta L_t = \frac{L_{it} - L_{(t-1)i}}{L_{(t-1)i}}$$

Keterangan :

ΔL_t = Perubahan Laba Perusahaan I pada tahun t

L_{it} = Laba perysahaan I pada tahun t

$L_{(t-1)i}$= Laba Perusahaan i pada tahun sebelumnya

2.9 *Rasio Keuangan*

Rasio dalam analisis laporan keuangan adalah angka yang menunjukan hubungan antara suatu unsur dengan unsur yang lainnya dalam laporan keuangan. Rasio merupakan alat untuk menyediakan pandangan terhadap kondisi yang mendasari. Menurut Chen *et al.* (1981), rasio keuangan telah memainkan peran penting sebagai alat evaluasi terhadap kemampuan dan kondisi keuangan perusahaan. Analisis rasio dapat mengungkapkan hubungan penting dan menjadi dasar perbandingan dalam menemukan kondisi dan tren yang sulit dideteksi dengan mempelajari masing-masing komponen yang membentuk rasio (Wild *et al.*, 2005).

Ada berbagai pendapat tentang kategori rasio berdasarkan tujuan penganalisis dalam mengevaluasi suatu perusahaan. Hanafi (2003) dan Ang (1997) mengelompokkan rasio keuangan sebagai berikut:

1. Rasio Likuiditas

Yaitu suatu rasio yang mengukur kemampuan suatu perusahaan dalam memenuhi kewajiban jangka pendeknya. Yang termasuk ke dalam rasio ini adalah rasio lancar (*current ratio*) dan rasio cepat.

2. Rasio Solvabilitas

Yaitu rasio yang mengukur kemampuan suatu perusahaan dalam memenuhi kewajiban jangka panjangnya atau mengukur tingkat proteksi kreditor jangka panjang. Yang termasuk ke dalam rasio ini adalah: *debt to equity ratio*, *leverage ratio* (*total debt to total asset ratio*), *debt service ratio*, dan *time interest earned ratio.*

3. Rasio Aktivitas

Yaitu rasio yang mengukur sejauh mana efektivitas penggunaan aktiva dengan melihat tingkat aktivitas aset. Yang termasuk ke dalam rasio ini adalah: *inventory turnover*, *total asset turnover, receivable turnover*, dan *average collection periode.*

4. Rasio Profitabilitas

Yaitu rasio yang bertujuan mengukur efektivitas manajemen yang tercermin pada imbalan hasil dari investasi melalui kegiatan penjualan. Yang termasuk ke dalam rasio ini adalah: *gross profit margin* (GPM), *operating profit margin* (OPM), *net profit margin* (NPM), *return on investment* (ROI), *return on equity (ROE).*

5. Rasio Pasar

Yaitu rasio yang lazim dan yang khusus digunakan di pasar modal yang menggambarkan situasi atau keadaan prestasi perusahaan di pasar modal. Namun, tidak berarti rasio lainnya tidak dipakai. Yang termasuk ke dalam rasio ini adalah: *price earning ratio (PER) dan price book value (PBV).*