

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Lahan Kritis**

Lahan kritis adalah lahan yang sudah tidak berfungsi lagi sebagai pengatur media, pengatur tata air, unsur produksi pertanian, maupun unsur perlindungan alam dan lingkungannya. Lahan kritis merupakan satu lahan yang kondisi tanahnya telah mengalami atau dalam proses kerusakan fisik, kimia, atau biologi yang akhirnya membahayakan fungsi hidrologi, orologi, produksi pertanian, pemukiman, dan kehidupan sosial ekonomi di sekitar daerah pengaruhnya (Ishak & Apong, 2012).

#### **2.2 Faktor – faktor penyebab lahan kritis**

Faktor-faktor penyebab lahan kritis cukup beragam, meliputi degradasi sifat fisik, kimia, dan biologi tanah. Lahan yang termasuk ke dalam katagori kemunduran sifat fisik tanah, diantaranya adalah yang disebabkan oleh tumbukan butir- butir hujan atau erosi, pemadatan tanah akibat penggunaan alat-alat dan mesin pertanian atau proses eluviasi, banjir dan genangan. Sedangkan lahan kritis yang disebabkan oleh kemunduran sifat kimia, diantaranya yang disebabkan oleh proses penggaraman, pemasaman, dan pencemaran oleh bahan agrokimia, serta pengurasan unsur hara tanaman (Kurnia, Sutrisno, & Sungkana, 2007).

Menurut (Dephut, 2009) tingkat kekritisian suatu DAS ditunjukkan oleh menurunnya penutupan vegetasi permanen dan meluasnya lahan kritis sehingga menurunkan kemampuan DAS dalam menyimpan air yang berdampak pada meningkatnya frekuensi banjir, erosi dan penyebaran tanah longsor pada musim

hujan dan kekeringan pada musim kemarau. Faktor yang mengakibatkan terjadinya tingkat kekritisn suatu DAS antara lain:

### **2.2.1 Pertumbuhan Penduduk**

Pertumbuhan penduduk mengakibatkan meningkatnya kebutuhan manusia itu sendiri terutama kebutuhan akan lahan (Sutapa, 2010). Permintaan akan lahan tersebut terus bertambah, sedangkan lahan yang tersedia jumlahnya terbatas (Huzaini, 2013). Timbulnya permasalahan akan penurunan kualitas lingkungan nantinya akan mengganggu keseimbangan ekosistem. Hal tersebut dikarenakan penggunaan lahan yang tidak memperhatikan kemampuan lahan, daya dukung dan fungsinya (Pewista & Harini, 2010).

### **2.2.2 Perambahan Hutan**

Salah satu faktor terjadinya lahan kritis pada DAS adalah terjadinya kerusakan hutan, salah satu penyebabnya adalah adanya perambahan hutan secara besar-besaran yang dilakukan penduduk. Peningkatan jumlah penduduk bukan hanya mengalihfungsikan lahan pertanian saja melainkan bisa mengancam keberadaan hutan sendiri dikarenakan kebutuhan akan lahan semakin meningkat mengakibatkan dilakukannya pembukaan lahan hutan semakin meningkat pula bahkan hutan yang dibuka secara besar-besaran dan dijadikan lahan pertanian ataupun non pertanian untuk keberlangsungan hidupnya (Susilawati, 2009).

### **2.2.3 Gempa Bumi dan Perubahan Iklim**

Kerusakan hutan bukan hanya disebabkan oleh manusia itu sendiri melainkan faktor alam yang terjadi semisal gempa bumi. Pada beberapa tempat, gempa bumi dapat mengakibatkan perubahan kestabilan tanah akibat sering

terjadinya tanah longsor. Demikian juga dengan perubahan iklim yang berakibat pada perubahan intensitas curah hujan, distribusi erosivitas hujan, dan sifat hujan lainnya yang akhirnya berakibat pada semakin tingginya erosi tanah dan sering terjadinya bencana banjir (Susilawati, 2009).

#### **2.2.4 Parameter Lahan Kritis**

Kriteria kekritisan lahan menurut Perdirjen BPDAS PS Nomor. P. 4/V-Set/2013 yang menggolongkan lahan kritis menjadi lima kelompok yaitu tidak kritis, potensial kritis, agak kritis, kritis dan sangat kritis. Kriteria ini didasarkan pada variabel-variabel yang terdiri dari, kondisi tutupan vegetasi, kemiringan lereng, tingkat erosi, manajemen, dan produktivitas lahan. Penilaian lahan kritis ditentukan berdasarkan fungsi lahan yaitu:

##### **2.2.4.1 Tingkat Bahaya Erosi**

Tingkat bahaya erosi dapat dihitung dengan cara membandingkan tingkat erosi di suatu satuan lahan dan kedalaman tanah efektif pada satuan lahan. Dalam hal ini tingkat erosi dihitung dengan menghitung perkiraan rata-rata tanah hilang tahunan akibat erosi lapis dan alur yang dihitung dengan rumus Universal Soil Loss Equation (USLE) (Ramayanti et al., 2015). Meningkatnya laju erosi hingga melebihi laju pembentukan tanah akan mengakibatkan lapisan tanah menjadi tipis dan bahkan mungkin hilang dan tinggal batuan dasarnya. Kehilangan lapisan tanah berarti kehilangan potensi untuk produksi pertanian dan kehutanan dan bahkan produksi air (Hendro, Nahdi, Budiastuti, & Purnomo, n.d.).

#### 2.2.4.2 Kemiringan Lereng

Kemiringan lereng adalah perbandingan antara jarak vertikal suatu lahan beberapa satuan, diantaranya adalah dengan persen dan derajat. Data spasial kemiringan lereng dapat disusun dari hasil pengolahan data garis kontur dengan bersumber pada peta topografi atau peta rupa bumi (Sunartomo, 2011). Kemiringan lereng merupakan faktor lain yang mempengaruhi keadaan lahan suatu wilayah. Wilayah yang terletak di dataran tinggi pada umumnya didominasi oleh lahan dengan kemiringan lereng di atas 15%. Kondisi wilayah tersebut berpotensi mengalami erosi yang besar (Hendro et al., n.d.).

#### 2.2.4.3 Manajemen Lahan

Manajemen lahan merupakan salah satu kriteria yang dipergunakan untuk menilai lahan kritis di kawasan hutan lindung, yang dinilai berdasarkan kelengkapan aspek pengelolaan yang meliputi keberadaan tata batas kawasan, pengamanan dan pengawasan serta dilaksanakan atau tidaknya penyuluhan (Kurnia et al., 2007).

#### 2.2.4.4 Produktivitas

Produktivitas merupakan salah satu kriteria yang dipergunakan untuk menilai kekritisian lahan di kawasan budidaya pertanian, yang dinilai berdasarkan rasio terhadap produksi komoditi umum optimal pada pengelolaan tradisional. Sehingga terdapat deliniasi wilayah kerja sesuai dengan tingkat kekritisian lahan dengan memperhatikan fungsi kawasan yaitu kawasan hutan lindung, kawasan budidaya pertanian, kawasan lindung di luar kawasan hutan, kawasan hutan konservasi dan kawasan hutan produksi (Wibowo & Gintings, 2010).

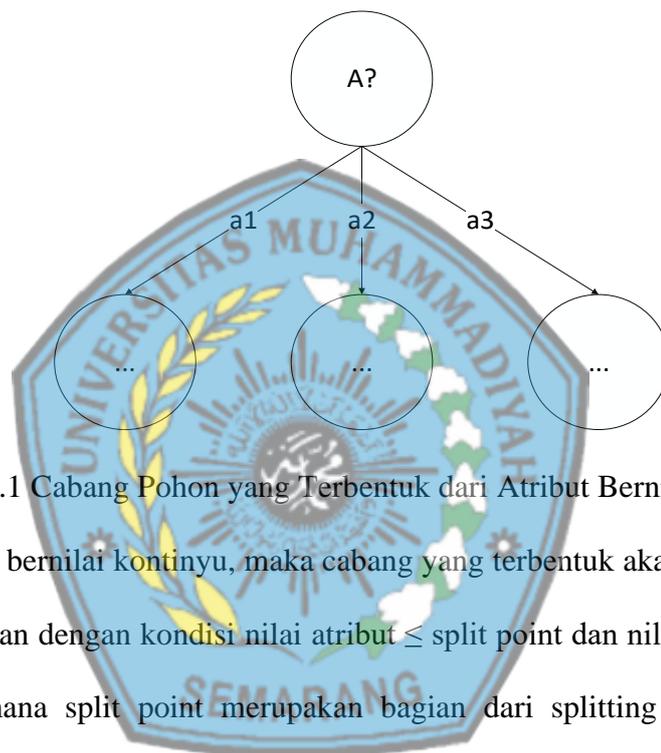
### 2.3 Algoritma C4.5

Algoritma C4.5 mirip sebuah struktur pohon dimana terdapat node internal yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas (Han et al., 2012). Algoritma C4.5 merupakan generasi baru dari algoritma ID3 (*iterative dichotomiser3*) yang dikembangkan oleh J. Ross Quinlan pada tahun 1983. Dalam ID3, induksi *decision tree* hanya dapat digunakan pada fitur bertipe kategorikal (nominal), sedangkan tipe numerik (ratio) tidak bisa digunakan (Singh & Gupta, 2014).

Algoritma C4.5 menggunakan konsep *information gain* atau *gain ratio* untuk menentukan *split* yang optimal (Larose, 2005). *Split* yang terpilih adalah yang mempunyai nilai *information gain* atau *gain ratio* yang terbesar. Tahapan dalam membentuk *Decision Tree* di C4.5 adalah (1) menghitung nilai *Entropy*, (2) menghitung nilai *Information Gain* atau *Gain Ratio* untuk masing-masing atribut, (3) atribut yang memiliki *Information Gain* atau *Gain Ratio* tertinggi dipilih menjadi akar (*root*) dan atribut yang memiliki nilai *Information Gain* atau *Gain Ratio* lebih rendah dari akar (*root*) dipilih menjadi cabang (*branches*), (4) menghitung lagi nilai *Information Gain* atau *Gain Ratio* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya, (5) atribut yang memiliki *Information Gain* atau *Gain Ratio* tertinggi dipilih menjadi cabang (*branches*), (6) mengulangi langkah ke-4 dan ke-5 sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa. Melalui tahapan ini akan terbentuk pohon (*tree*) yang terdiri dari akar (*root*), cabang (*branches*) dan daun (*leaf*) yang ditentukan melalui proses pemilihan *split* atribut.

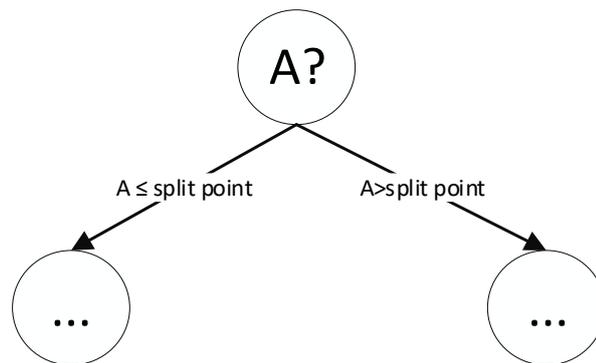
Dalam proses pengujian atribut, cabang baru yang terbentuk akan diperhatikan dari tipe atribut (Han et al., 2012). Berikut 3 jenis cabang (*branches*) yang mungkin muncul dalam pohon keputusan (*Decision Tree*).

1. Jika atribut bernilai diskret, maka cabang yang terbentuk akan selalu sama dengan jumlah variasi kasus yang terdapat pada atribut tersebut.



Gambar 1.1 Cabang Pohon yang Terbentuk dari Atribut Bernilai Diskret

2. Jika atribut bernilai kontinu, maka cabang yang terbentuk akan mempunyai 2 kemungkinan dengan kondisi nilai atribut  $\leq$  split point dan nilai atribut  $>$  split point. Dimana split point merupakan bagian dari splitting criterion. Data kontinu adalah data yang dapat digunakan untuk operasi hitung, data yang diperoleh dari hasil perhitungan atau pengukuran (Sugianto, 2016). Contoh dari data ini adalah jumlah benar atau salah dalam suatu tes, skor nilai, ranking, tinggi badan, berat badan, panjang, jarak dll.



Gambar 1.2 Cabang Pohon yang Terbentuk dari Atribut Bernilai Kontinyu

Perhitungan *split point* pada gambar 2.1 didapat dari persamaan

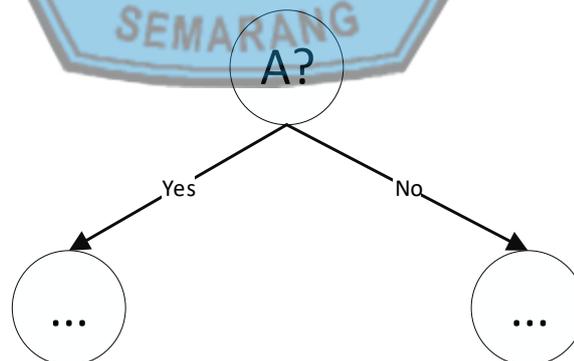
$$\text{Split Point} = \frac{a_1 + a_{i+1}}{2} \quad (1.1)$$

Berdasarkan persamaan di atas, berikut merupakan keterangan:

$a_1$  = nilai saat ini

$a_{i+1}$  = nilai selanjutnya

3. Jika atribut bernilai diskret dan bernilai biner, maka cabang yang terbentuk akan selalu dua dengan nilai Yes atau No.



Gambar 1.3 Cabang Pohon yang Terbentuk dari Atribut Bernilai Diskret dan Biner

## 2.4 Pemilihan Split Atribut

Pemilihan split atribut adalah proses yang fundamental di dalam proses induksi *Decision Tree*. Dalam setiap proses induksi dalam membentuk *Decision Tree*, pemilihan split atribut sangat penting dikarenakan sebagai penentu apakah sebuah atribut akan di-split atau tidak. Ada dua metode pemilihan split atribut paling populer yaitu *information gain* dan *gain ratio* (Han, Kamber, & Pei, 2012).

### 2.4.1 Metode *Information Gain*

Metode penelitian ini diperkenalkan oleh Quinlan dengan berdasar model ID3 (*Iterative Dichotomiser 3*). Metode yang diperkenalkan Quinlan cocok untuk *dataset* dengan variabel diskret akan tetapi metode yang diperkenalkan tidak cocok untuk *dataset* dengan *missing value*. Metode penelitian Quinlan menggunakan pemilihan *split* atribut yang disebut *Gain*. Sebelum mencari nilai *gain*, hitung nilai *Entropy*. *Entropy* didefinisikan sebagai nilai informasi yang diharapkan. Nilai *Entropy* dapat dihitung melalui rumus persamaan dibawah ini:

$$Entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (1.2)$$

Berdasarkan persamaan diatas, berikut merupakan keterangannya:

S = himpunan kasus

n = jumlah partisi S

$p_i$  = proporsi dari  $S_i$  terhadap S

Perhitungan nilai *Gain* dapat dirumuskan dengan persamaan dibawah ini:

$$Gain(S, A) = Entropy(s) - \sum_i^m \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (1.3)$$

Berdasarkan persamaan diatas, berikut merupakan keterangannya:

S: himpunan dataset

A: kriteria atau atribut

m: jumlah nilai yang mungkin pada kriteria A (jumlah kelas)

$S_i$ : himpunan dataset untuk nilai m

$|S_i|$ : jumlah dataset untuk nilai m

$|S|$ : jumlah dataset dalam S

Kemudian dipilih atribut yang mempunyai nilai *information gain* yang terbesar.

#### 2.4.2 Metode Gain Ratio

Metode penelitian ini merupakan pengembangan dari metode *Iterative Dichotomiser 3* (ID3). Quinlan memperkenalkan metode ini untuk pemilihan *split* atribut menggunakan metode *Info Gain Ratio* (IGR) memperbaiki *Information Gain* (IG). C4.5 yang diperkenalkan dapat bekerja pada variabel kontinu dan *missing value*.

Rumus persamaan *Gain Ratio* seperti berikut:

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Info}(S, A)} \quad (1.4)$$

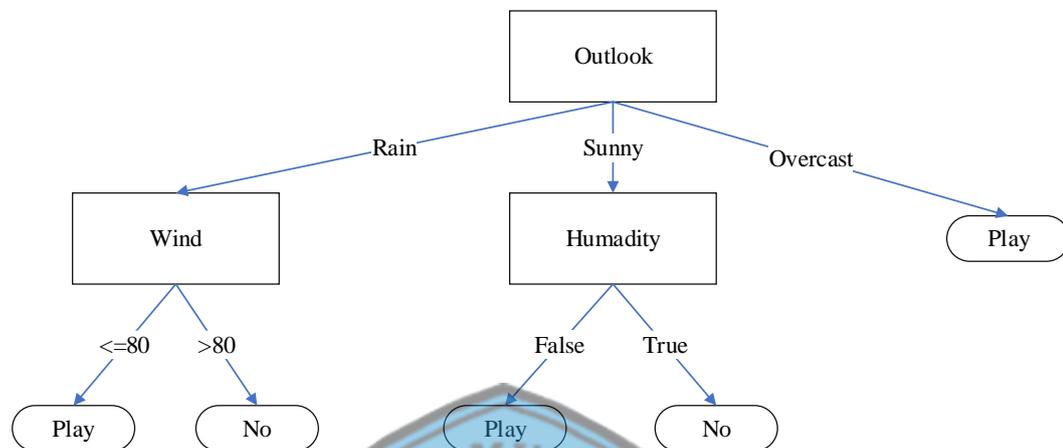
Dimana *split info* mempunyai persamaan rumus sebagai berikut:

$$\text{Split Information}(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (1.5)$$

#### 2.5 Rule Set

*Decision Tree* dapat sulit dimengerti karena informasi dari sebuah *class* tersebar di seluruh pohon (*tree*). C4.5 memperkenalkan bentuk lain yang terdiri dari daftar *rule* yang berbentuk “Jika A dan B dan C. maka class X”(Wu et al., 2008). Rule set C4.5

terbentuk dari *initial tree* yang belum *di-pruning*. Dimana tiap bagian dari mulai akar (*root*) sampai daun (*leaf*) merupakan *prototipe* dari *rule set*.



Gambar 1.4 Contoh Pohon Keputusan (*Decision Tree*)

*Rule set* dari Gambar 2.4 dapat diterjemahkan sebagai berikut:

*Rule 1: IF Outlook = Rain AND Wind <= 80 THEN Play*

*Rule 2: IF Outlook = Rain AND Wind > 80 THEN No*

*Rule 3: IF Outlook = Sunny AND Humidity = False THEN Play*

*Rule 4: IF Outlook = Sunny AND Humidity = True THEN No*

*Rule 5: IF Outlook = Overcast THEN Play*

## 2.6 Evaluasi dan Validasi

### 2.6.1 10-Fold Cross Validation

*Cross validation* adalah metode statistik yang penting dengan berbagai variasi bentuk berdasar basis tertentu (Witten, Frank, & Hall, 2011). Model validasi klasifikasi yang sering digunakan adalah *k-fold cross validation* (Floudas, 2009). *K-Fold Cross validation* adalah teknik validasi dengan membagi data secara acak ke dalam *k* bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Han et al., 2012). *Dataset* dibagi menjadi *m* bagian (*subsets*) dengan ukuran yang

relatif sama. Model klasifikasi akan ditraining sebanyak  $m$  kali, dan terdapat 1 bagian (*subset*) yang disembunyikan selama *testing* untuk tiap-tiap *split*. *10-fold cross validation* merupakan *cross validation* yang direkomendasikan untuk perhitungan akurasi (Torgo, 2011). *Dataset* dibagi menjadi 10 bagian (*subsets*). 1 *split* terdiri dari data test dan data training. Tiap *split* ditraining dan ditest dan proses ini diulang sampai dengan 10 kali.

DATASET									
Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Split 8	Split 9	Split 10
Training									Test
Training								Test	
Training							Test		
Training						Test			
Training					Test				Training
Training				Test				Training	
Training			Test				Training		
Training		Test					Training		
Training	Test						Training		

Gambar 1.5 Pembagian Dataset untuk 10-Fold Cross Validation

### 2.6.2 Confusion Matrix

Akurasi pengklasifikasian dari model yang dievaluasi pada *m-fold cross validation* dapat direpresentasikan menggunakan matrik klasifikasi atau *confusion matrix* (Floudas, 2009). *Confusion matrix* merupakan tabulasi dari evaluasi performa model pengklasifikasian berdasarkan perhitungan objek dengan nilai benar dan salah pada model prediksi (Gorunescu, 2011). *Confusion matrix* merupakan matrik 2 dimensi yang menggambarkan perbandingan antara nilai prediksi dengan nilai kenyataan, ditunjukkan pada Tabel 2.1. Jika nilai prediksi benar dan nilai kenyataannya benar, maka disebut *True Positive* (TP). Jika nilai prediksi benar dan nilai kenyataannya salah, maka disebut *False Positive* (FP). Jika

nilai prediksi salah dan nilai kenyataannya benar, maka disebut *False Negative* (FN). Jika nilai prediksi salah dan nilai kenyataannya salah maka disebut *True Negative* (TN).

Tabel 1.1 *Confusion Matrix*

<i>Predicted Class</i>	<i>Actual Class</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Setelah mendapatkan *confusion matrix*, dilanjutkan dengan perhitungan kinerja pengklasifikasi. Beberapa perhitungan yang umum digunakan untuk memahami dan menjelaskan kinerja pengklasifikasi *accuracy*, *precision*, *sensitivity*, dan *specificity* (Vercellis, 2009).

Sensitivitas adalah kemampuan pengklasifikasi untuk memilih semua kasus yang harus dipilih. Sensitivitas disebut juga *recall* atau *True Positive Rate* (TPrate). Persamaan berikut digunakan untuk menghitung sensitivitas.

$$\text{Sensitivity} = \text{recall} = \frac{TP}{TP+FN} \quad (1.6)$$

Perhitungan sensitivitas saja tidak cocok untuk mengevaluasi sebuah pengklasifikasi (Kotu & Deshpande, 2015), dibutuhkan pengukuran lanjutan yaitu *specificity*.

*Specificity* adalah kemampuan pengklasifikasi untuk menolak semua kasus yang perlu ditolak. *Specificity* atau disebut *True Negative Rate* (TNrate). *Specificity* mempunyai beberapa nilai *False Positive* (FP). Persamaan 2.6 digunakan untuk menghitung *specificity*.

$$Specificity = \frac{TN}{TN+FN} = 1 - sensitivity \quad (1.7)$$

Akurasi didefinisikan sebagai kemampuan pengklasifikasi untuk memilih semua kasus yang harus dipilih dan menolak semua kasus yang harus ditolak. Jika sebuah pengklasifikasi mempunyai akurasi 100% berarti pengklasifikasi tersebut mempunyai False Negative (FN) = False Positive (FP) = 0. Akurasi dihitung menggunakan Persamaan 2.8.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1.8)$$

*Precision* digunakan untuk mengukur berapa persen ketepatan prediksi kelas dari suatu data, yaitu berapa banyak kelas yang diprediksi positive dan sesuai dengan kenyataannya.

$$Precision = \frac{TP}{TP+FP} \quad (1.9)$$

Rata-rata tertimbang atau weighted average digunakan untuk menghitung nilai rata-rata suatu kelas sesuai dengan jumlah instance yang dimiliki.

$$\bar{x} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n} \quad (1.10)$$

### 2.6.3 F-Measure

F-Measure merupakan salah satu perhitungan evaluasi yang perhitungannya mengharmonikan rata-rata dari *precision* dan *recall* (Venkatesan, 2014). Sehingga menghasilkan metrik yang efektif untuk pencarian kembali informasi dalam himpunan yang mengandung masalah ketidakseimbangan. Nilai sensitivitas dan presisi pada suatu keadaan dapat memiliki bobot yang berbeda. Nilai F-Measure akan meningkat jika nilai dari *precision* dan *recall* juga meningkat. Nilai F-Measure

yang tinggi menunjukkan bahwa model tersebut bekerja lebih baik pada kelas positif (Bekkar, Djemaa, & Alitouche, 2013).

$$F - Measure = \frac{2 \times sens \times prec}{sens + prec} \quad (1.11)$$

