

**ESTIMASI LUAS AREAL HUTAN  
MENGUNAKAN KOMBINASI ALGORITMA GENETIKA  
DAN SUPPORT VECTOR MACHINE**

***THE ESTIMATION OF THE FOREST AREA  
USING THE COMBINATION OF GENETIC ALGORITHM  
AND THE SUPPORT VECTOR MACHINE***

**LAPU' TOMBILAYUK**



**PROGRAM PASCASARJANA  
UNIVERSITAS HASANUDDIN  
MAKASSAR**

**2013**

## ABSTRAK

**LAPU TOMBILAYUK.** Estimasi Luas Areal Hutan Menggunakan Kombinasi Algoritma Genetika Dan *Support Vector Machine* (dibimbing oleh **Syafaruddin dan Loeky Haryanto**)

Penelitian ini bertujuan membuat sebuah model estimasi luas areal hutan menggunakan algoritma genetika yang dikombinasikan dengan algoritma *support vector machine*.

Penelitian ini menggunakan data kondisi hutan Taman Nasional Kutai (TNK), Kalimantan Timur, tahun 2003-2012 dengan total luas areal sebesar 198.629 hektar sebagai objek luas areal hutan yang diestimasi. Data kondisi hutan tersebut diklasifikasikan menggunakan algoritma *support vector machine* terhadap pemicu yang mengurangi luas areal hutan dan pemicu yang menambah luas areal hutan. Sementara algoritma genetika digunakan untuk optimalisasi proses estimasi luas areal hutan. Setiap generasi yang diperoleh dari proses algoritma genetika merupakan estimasi kondisi luas areal hutan pada tahun berikutnya.

Hasil penelitian menunjukkan adanya tren estimasi yang meningkat secara fluktuatif. Pertambahan kerusakan hutan tidak meningkat secara linier setiap tahunnya. Hal ini menggambarkan bahwa kombinasi algoritma genetika dan *support vector machine* bekerja secara rasional terhadap variabel yang mengurangi luas areal hutan dan variabel yang menambah luas areal hutan. Luas areal hutan sebesar 198.629 hektar diestimasi akan habis sekitar 153 tahun akan datang.

Kata Kunci : *estimasi, algoritma genetika, support vector machine.*

## ABSTRACT

**LAPU TOMBILAYUK.** *The Estimation of the Forest Areas Using The Combination of the Genetic Algorithm and the Support Vector Machine* (supervised by **Syafaruddin and Loeky Haryanto**)

This research aimed to create a model to estimate the forest areas by using the Genetic Algorithm combined with the Support Vector Machine.

The data of the forest areas of about 198,629 hectares were the data of the forest conditions in Kutai National Park (KNP), East Kalimantan during the period between 2003 and 2012. The above data were then classified using algorithm of the support vector machine in order to find out what factors had triggered the reduction of the forest areas, and what factors had triggered the addition of the forest areas. After that, the genetic algorithm was used to optimize the process of estimating the forest areas. Each generation obtained from the genetic algorithm was considered as an estimation of the forest area condition for the following years.

The research result indicated that the estimation tended to fluctuate. The damaged forest did not increase in a linear way every year. This illustrated that the combination between the genetic algorithm and the support vector machine operated rationally on the variables reducing and increasing the forest area. Thus, it was estimated that the forest areas of 198,629 hectares would all be gone in about 153 years' time.

*Keywords: Estimation, Genetic Algorithm, Support Vector Machine.*

## KATA PENGANTAR

Ucapan syukur bagi TUHAN Pencipta Alam Semesta, Ilmu dan Pengetahuan, sungguh besar penyertaan-Nya yang memberikan kemampuan bagi penulis sehingga dapat menyelesaikan penulisan tesis ini dengan judul “***Estimasi Luas Areal Hutan Menggunakan Kombinasi Algoritma Genetika dan Support Vector Machine***”.

Gagasan yang melatarbelakangi penelitian ini adalah bagaimana menerapkan algoritma genetika dikombinasikan dengan algoritma support vector machine untuk dapat meramalkan atau mengestimasi luas areal hutan Taman Nasional Kutai (TNK) Kalimantan Timur dimasa yang akan datang dengan melihat kondisi sekarang dan kondisi yang masa lalu.

Banyak kendala yang dihadapi oleh penulis dalam rangka penyusunan tesis ini, namun berkat bantuan dari berbagai pihak, maka tesis ini akhirnya dapat terselesaikan. Penulis dengan tulus menyampaikan ucapan terima kasih kepada :

1. Prof. Dr. Ir. H. Salama Manjang, MT., selaku Ketua Program Studi Teknik Elektro
2. **Dr. Eng. Syafaruddin, ST., M.Eng**, selaku Ketua Komisi Penasehat, atas bantuan dan bimbingan yang telah diberikan mulai dari awal penyusunan proposal hingga selesainya penelitian ini.
3. **Dr. Loeky Haryanto, MS., M.A.T.**, sebagai Anggota Komisi Penasehat atas bantuan dan bimbingan yang telah diberikan mulai dari awal penyusunan proposal hingga selesainya penelitian ini.

4. DR. Ir. H. Rhiza S. Sadjad, MSEE, Dr. Ir. H. Andani Achmad, MT., Dr. Elyas Palentei, ST., M.Eng., selaku Dosen Penguji yang telah memberikan saran dalam perbaikan tesis ini.
5. Dr. Ir. Erly Sukrismanto, M.Sc., selaku Kepala Balai Taman Nasional Kutai (TNK) Kalimantan Timur yang telah memberikan izin pengambilan data dilokasi dan izin masuk kawasan konservasi (SIMAKSI).
6. Zahrotun Nizaa' selaku staf Urusan Pelayanan Balai Taman Nasional Kutai (TNK) Kalimantan Timur, yang telah membantu penulis mendapatkan data deforestasi hutan Taman Nasional Kutai.
7. Istriku tercinta Marcelina Biantong, anakku tercinta Sendy Yeremia Rita, Efraim Biantong Rita, Valerian Tombilayuk, kedua orang tua, beserta keluarga yang tiada henti berdoa dan berharap akan keberhasilan penulis.
8. Terima kasih juga penulis sampaikan kepada teman-teman Pasca Sarjana UNHAS, Fakultas Teknik Elektro, Program Studi Teknik Informatika, angkatan 2011 yang telah membantu dan mendoakan penulis dalam menyelesaikan penelitian ini.
9. Wilem Musu, S.P, S.Kom, MT., selaku sahabat Pasca Sarjana UNHAS, Program Studi Teknik Informatika angkatan 2011 yang telah mendoakan dan membantu penulis dalam menyelesaikan penelitian ini.

Kami menyadari keterbatasan dan kelemahan kami dalam penelitian ini, walaupun telah dilakukan usaha yang maksimal. Oleh karena itu diharapkan saran dan kritik membangun untuk kesempurnaan penulisan thesis ini, dan untuk pengembangan penelitian di masa yang akan datang.

Akhir kata, penulis berharap semoga tesis ini dapat memberikan manfaat dan mendapatkan rahmat dari Tuhan Yang Maha Esa.

Makassar, 1 Agustus 2013

P e n u l i s,

**( Lapu' Tombilayuk )**

# **BAB I**

## **PENDAHULUAN**

### **I.1. Latar Belakang**

Kawasan hutan di Indonesia mengalami kerusakan (*deforestasi*) yang terus meningkat setiap tahunnya dimana deforestasi terbesar terjadi di Kalimantan dan Sumatera. Kerusakan utama disebabkan oleh pengelolaan hutan yang tidak tepat, industri kayu semakin banyak, penebangan liar (*illegal logging*), *kebakaran*, pengalihan fungsi (konversi) hutan untuk perkebunan rakyat, serta pembukaan lahan tambang. Dampak dari kerusakan tersebut mengakibatkan berbagai bencana ekologi seperti banjir, longsor, kekeringan dan pemanasan global (*global warming*) yang semakin meningkat setiap tahunnya [1].

Pemerintah melalui Menteri Kehutanan mengeluarkan kebijakan penebangan hutan alam primer dan hutan gambut tidak boleh dilakukan lagi seiring upaya yang dilakukan Kementerian kehutanan untuk memulihkan kondisi hutan Indonesia. Sementara itu perizinan hutan lindung yang digunakan sebagai wilayah pertambangan terbuka tidak akan dilakukan kembali. Peraturan Pemerintah Republik Indonesia Nomor 10 Tahun 2010 Tentang Tata Cara Perubahan Peruntukan dan Fungsi Kawasan Hutan, mengatur bagaimana perubahan kawasan khususnya fungsi kawasan [ 2].

Walaupun kebijakan pemerintah tersebut telah dituangkan dalam bentuk Peraturan Pemerintah (PP) dan undang-undang, tetapi deforestasi

masih saja terus terjadi. Bahkan Forest Watch Indonesia (FWI) sebuah jaringan pemantau hutan independen Indonesia memproyeksikan pada tahun 2020 hutan di pulau Jawa akan habis, Bali dan Nusa Tenggara tersisa 0,08 juta ha, Sulawesi 7,20 juta ha, Sumatra 7,72 juta ha dan Kalimantan 21,29 juta ha serta Papua 33,45 juta ha. Dalam kurun waktu 2000-2009 tercatat laju deforestasi sebesar 1,51 juta ha hutan per tahun dimana deforestasi terbesar terjadi di Kalimantan sebesar 0,55 juta ha per tahun [3].

Banyak penelitian yang dilakukan untuk menghambat laju deforestasi yang terus meningkat. Salah satu penelitian yang telah dilakukan adalah memprediksi tingkat kelestarian dan produksi kayu di KPH (Kesatuan Pengelolaan Hutan) Bojonegoro yang dilakukan oleh IPB [4]. Untuk memprediksi kelestarian dan produksi kayu tersebut digunakan data selama 10 tahun terakhir. Data dianalisis dengan cara memproyeksikan melalui persentase perubahan/dinamika populasi dari suatu periode ke periode berikutnya berdasarkan luas lahan sebelumnya dan luas lahan sesudahnya. Presentasi dinamika populasi ini dihitung untuk variabel penyebab kerusakan dan variabel pelestarian hutan dengan mengasumsikan kondisi minimum dan maksimum dari tiap-tiap variabel.

Salah satu metode yang akan digunakan dalam penelitian ini untuk memprediksi atau mengestimasi areal hutan dalam Taman Nasional Kutai (TNK) Kalimantan Timur adalah Algoritma Genetika atau *Genetic Algorithm* (GA) dikombinasikan dengan algoritma Support Vector Machine

(SVM). Algoritma Genetika adalah Teknik pencarian (*heuristic*) yang didasarkan pada gagasan evolusi seleksi alam. Algoritma ini memanfaatkan proses seleksi alamiah yang dikenal dengan proses evolusi. Dalam proses evolusi, individu secara terus menerus mengalami perubahan gen untuk menyesuaikan dengan lingkungan hidupnya (*survival of the fittest*). Proses alamiah ini melibatkan perubahan gen yang terjadi pada individu melalui proses perkembangbiakan.

*Support Vector Machine* (SVM) adalah suatu metode pengenalan pola (*pattern recognition*) yang merupakan suatu bidang dalam komputer sains, yang memetakan suatu data ke dalam suatu konsep tertentu yang telah didefinisikan sebelumnya. Konsep tertentu ini disebut class atau category. Metode ini banyak diaplikasikan dalam bioinformatika, seperti mengenali suara dalam sistem sekuriti, membaca huruf dalam optical character recognition (OCR), mengklasifikasikan penyakit secara otomatis berdasarkan hasil kondisi medis pasien, serta metode dapat digunakan untuk meramalkan atau memprediksi kemungkinan-kemungkinan yang terjadi di kemudian hari berdasarkan gejala atau kejadian sebelumnya.

Beberapa peneliti sebelumnya menggabungkan metode algoritma genetika (*genetic Algorithm*) dengan support vector machine (SVM) untuk mendapatkan hasil ramalan terbaik, algoritma genetika berperan untuk mengoptimasi nilai parameter *support vector machine*. Paid an Hong (2005) telah melakukan peramalan beban listrik regional berdasarkan recurrent support vector machine dengan algoritma genetika, adapula

penelitian yang dilakukan oleh Lessmann, Stahlbock, Crone pada tahun 2006 tentang algoritma genetika untuk pemilihan model support vector machine , dan aplikasi *genetic algorithm - support vector machine* (GA-SVM) untuk prediksi aktivitas saluran kalium, yang diteliti oleh Pourbasheer, Rihai, Ganjali, dan Norouzi pada tahun 2009). Ada dua penelitian yang terbaru dengan menerapkan algoritma genetika dikombinasikan dengan *support vector machine*, yang pertama adalah peramalan harga rumah oleh Jirong Gu, Mingcang Zhu, dan Liungaungyan Jiang pada tahun 2011, yang kedua adalah penerapan metode algoritma genetika dengan support vector machine untuk meramalkan beban listrik di PT. PLN (Persero) sub unit penyaluran dan pusat pengatur beban (P3B) Jawa Timur-Bali, dilakukan oleh Ni Luh Puteh Satyaning Pradnya Paramita, pada tahun 2012 [5].

Penelitian ini menerapkan metode algoritma genetika dengan support vector machine untuk mengestimasi luas areal hutan Taman Nasional Kutai Kalimantan Timur. Algoritma genetika dan support vector machine digunakan karena metode ini dapat menganalisis data *time series* (serangkaian nilai-nilai variabel yang disusun berdasarkan waktu) untuk mengetahui perkembangan suatu atau beberapa kejadian dan pengaruhnya pada kejadian lain. Dari beberapa peneliti sebelumnya menyimpulkan bahwa metode algoritma genetika jika digabungkan dengan support vector machine akan menghasilkan peramalan yang

akurat dengan menggunakan pengujian *Mean Absolut Percentage Error (MAPE)* yaitu nilai tengah kesalahan persentase absolute.

Beberapa metode peramalan yang bisa digunakan untuk mengestimasi luas areal hutan, yaitu metode parametric (merumuskan model matematika atau statistic) yang telah digunakan oleh Desi Yuniarti untuk meramalkan jumlah penumpang yang berangkat melalui Bandar Udara Temindung Samarinda tahun 2012, dengan pendekatan metode *Autoregressive Integrated Moving Average (ARIMA)* yang dikembangkan oleh Box Jenkins, diperoleh hasil peramalan konstan dari Januari 2012 sampai Desember 2012 yaitu sebesar 4035 penumpang [6]. Kelebihan dari metode ini adalah dapat memberikan perhitungan yang lebih tepat, sedangkan kelemahan adalah tidak dapat digunakan untuk meramalkan data yang banyak dan bervariasi.

Jika metode ini diterapkan untuk mengestimasi luas lahan hutan TNK, maka hanya bisa mengolah data pemicu pengurang lahan hutan (kebakaran, perambahan dan penebangan liar), dan tidak dapat mengolah data bervariasi yaitu data penambah luas areal hutan (reboisasi) dan data pengurang luas areal hutan. Dan jika menerapkan algoritma genetika dengan support vector machine, maka data-data tersebut dapat dianalisis.

Metode nonparametrik (memungkinkan ramalan akan dihitung langsung dari data historis), telah digunakan oleh Melly Lukman dan Eko Susanto tahun 2007 menggunakan teknik *exponensial smoothing* untuk peramalan debit aliran sungai. Hasil ramalan adalah nilai debit rata-rata

yang sama untuk setiap bulan (Januari-Desember 2002) yaitu 85.32 m<sup>3</sup>/detik [7]. Kelebihan metode ini adalah lebih baik untuk peramalan jangka pendek dan baik digunakan bila tidak ada trend atau komponen musiman, kelemahan metode ini adalah hasil yang didapatkan belum optimal dan akurat meramal.

Jika metode ini diterapkan untuk mengestimasi luas areal hutan TNK, maka masa waktu yang akan diprediksi sangat singkat (kurang dari 10 tahun), dan tidak optimal dalam pencarian nilai parameter. Algoritma genetika dapat mengoptimalkan nilai parameter support vector machine, sehingga waktu estimasi luas areal hutan dalam jangka waktu yang lama (diatas sepuluh tahun).

Metode berbasis kecerdasan buatan seperti *artificial neural network* (ANN) telah digunakan oleh I Made Mataram untuk meramalkan beban hari libur, pada tahun 2008, hasil penelitian adalah dapat meramalkan beban harian selama 24 jam dengan *error* rata-rata yang sangat kecil, yaitu tahun 2006 didapatkan *error* rata-rata di bawah 1% [8]. Metode ini memiliki kemampuan untuk mempelajari dan mengadaptasi situasi baru dengan mengingat pola data sebelumnya, kekurangan jika metode ini diterapkan pada analisis data runtun waktu (*time series*) akan sulit dilakukan peramalan.

Jika metode ini diterapkan untuk mengestimasi luas areal hutan TNK, maka hasil ramalan yang diperoleh tidak tepat dan tidak bisa mencari

pemicu-pemicu mana yang lebih banyak menyebabkan berkurangnya areal hutan.

Penerapan algoritma support vector machine, dapat memisahkan data-data pemicu berdasarkan luas lahan terbesar dan terkecil.

## **I.2. Rumusan Masalah**

Dari latar belakang tersebut, maka beberapa permasalahan yang akan menjadi rumusan adalah sebagai berikut :

1. Bagaimana mengestimasi areal hutan Taman Nasional Kutai (TNK).
2. Bagaimana menerapkan algoritma genetika dan support vector machine untuk mengestimasi areal hutan Taman Nasional Kutai.

## **I.3. Tujuan Penelitian**

Tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut :

1. Menghasilkan aplikasi yang dapat mengestimasi areal hutan Taman Nasional Kutai.
2. Penerapan algoritma genetika dan support vector machine untuk menyelesaikan permasalahan estimasi.

## **I.4. Manfaat Penelitian**

1. Untuk mengetahui areal hutan Taman Nasional (TNK) yang rusak, dengan penerapan metode algoritma genetika dan support vector machine.
2. Sebagai bahan rekomendasi untuk Balai Taman Nasional Kutai Kalimantan Timur dalam penggunaan metode terbaik yang dapat digunakan untuk mengestimasi areal hutan Taman Nasional Kutai, di waktu sekarang dan akan datang.

### **I.5. Batasan Masalah**

Untuk mencapai sasaran dan tujuan dari penelitian ini, maka permasalahan yang ada akan dibatasi pada :

1. Luas wilayah cakupan hutan pada Taman Nasional Kutai (TNK) Kalimantan Timur.
2. Data yang dipakai adalah data pemicu kerusakan areal hutan Taman Nasional Kutai Kalimantan Timur, dari tahun 2003 sampai dengan tahun 2012.
3. Bahasa pemrograman yang digunakan untuk mengolah data dengan metode algoritma genetika dan support vector machine adalah Borland Delphi versi 7.0

### **I.6. Sistematika Penulisan**

#### **BAB I. PENDAHULUAN**

Pada BAB ini disajikan Latar Belakang Masalah, Rumusan Masalah, Tujuan Penelitian, Manfaat Penelitian, Batasan Masalah dan Sistematika Penulisan.

#### **BAB II. TINJAUAN PUSTAKA**

Pada BAB ini disajikan tentang teori-teori yang berhubungan dengan fakta atau kasus yang sedang dibahas. Disamping itu juga dapat disajikan teori-teori metode yang diterapkan dalam penelitian ini, serta road map penelitian yang menjelaskan peneliti-peneliti yang sudah menerapkan metode-metode yang diterapkan dalam penelitian ini.

### BAB III. METODE PENELITIAN

Pada BAB ini disajikan waktu dan lokasi penelitian, metode pendekatan, tahap penelitian, instrument penelitian, perancangan sistem, dan kerangka berpikir.

### BAB IV. HASIL PENELITIAN DAN PEMBAHASAN

Pada BAB ini disajikan perancangan aplikasi, yang terdiri dari perancangan *Unified Modeling Language* (UML), perancangan DataBase, perancangan antar muka (*User Interface*), pengujian akurasi prediksi, dan pengujian *White Box*.

### BAB V. KESIMPULAN DAN SARAN

BAB ini menyatakan penyimpulan dari semua yang telah dicapai di pada masing-masing BAB sebelumnya. Tersusun atas Kesimpulan dan Saran.

## **BAB II**

### **TINJAUAN PUSTAKA**

Peramalan atau estimasi yang tepat atas suatu kejadian di masa yang akan datang, sangat diperlukan dalam pengambilan keputusan untuk penyelesaian suatu masalah. Seringkali ramalan yang dilakukan seseorang atau pimpinan suatu organisasi tidak tepat, disebabkan karena berdasarkan data yang tidak benar atau berdasarkan logika hayalan dan juga penggunaan metode ramalan secara manual. Beberapa peneliti sebelumnya menggunakan algoritma genetika dan algoritma support vector machine untuk memprediksi kejadian yang akan terjadi di masa yang akan datang. Pada penelitian ini akan merancang aplikasi perangkat lunak dengan menerapkan pendekatan Algoritma Genetika dikombinasikan dengan Algoritma Support Vector Machine (SVM) untuk meramalkan atau estimasi luas areal hutan Taman Nasional Kutai Kalimantan Timur, di masa yang akan datang. Dari aplikasi ini diharapkan menghasilkan estimasi dengan nilai akurasi lebih baik.

#### **II.1. Algoritma Genetika**

Algoritma Genetika atau *Genetic Algorithm* (GA) muncul dari inspirasi oleh teori Charles Darwin tentang prinsip genetika dan seleksi alam (teori evolusi) yang ditemukan oleh John Holland dari Universitas Michigan, Amerika Serikat pada tahun 1975, melalui sebuah penelitian dan

dipopulerkan oleh salah satu muridnya, yaitu David Goldberg. Algoritma genetika telah diaplikasikan oleh beberapa peneliti sebelumnya untuk memecahkan masalah optimasi pada berbagai bidang permasalahan teknik, bisnis, dan hiburan. Konsep dasar algoritma genetika dirancang untuk mensimulasikan proses-proses dalam sistem alam yang diperlukan untuk evolusi (perubahan secara berangsur-angsur), khususnya teori evolusi. Konsep ini menurut Charles Darwin bahwa teori evolusi adalah di alam terjadi persaingan antara individu-individu untuk memperebutkan sumber daya alam yang langka sehingga makhluk yang kuat mendominasi atau mengalahkan yang lebih lemah (*survival of the fittest*) (T.Sutojo, dkk., 2011) [9].

Dalam algoritma genetika dikenal istilah-istilah sebagai berikut :

- 1) Gen (*Genotype*) adalah variabel dasar yang membentuk sebuah nilai tertentu dalam satu kesatuan gen yang dinamakan kromosom. Dalam algoritma genetika, gen ini bisa berupa *biner*, *float*, *integer* maupun karakter, atau kombinatorial
- 2) *Allele* adalah nilai dari suatu gen, bisa berupa biner, float (bilangan real), integer, maupun karakter.
- 3) Kromosom atau Individu adalah gabungan dari gen-gen yang membentuk arti tertentu. Ada beberapa bentuk kromosom, yaitu :
  - Kromosom biner , adalah kromosom yang disusun dari gen-gen yang bernilai biner. Jumlah gen pada kromosom biner menunjukkan tingkat

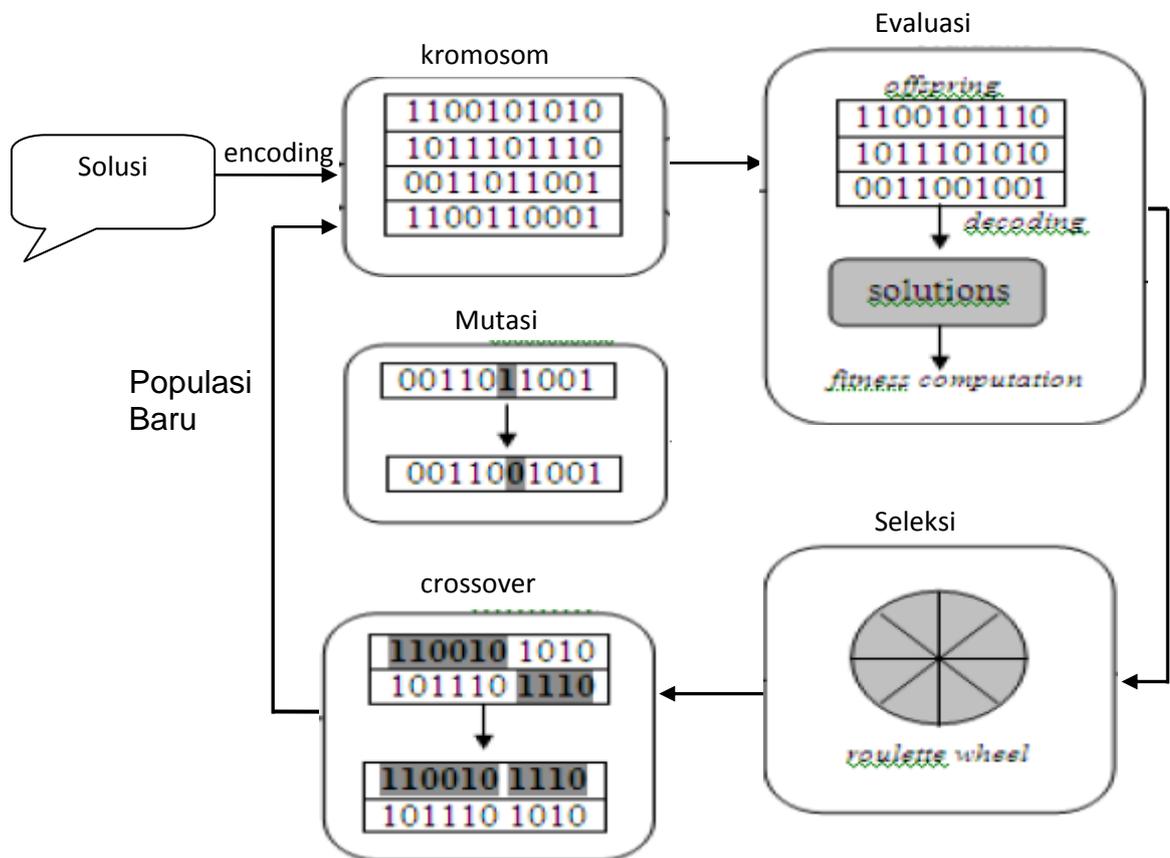
ketelitian yang diharapkan. Kromosom ini bagus bila digunakan untuk permasalahan yang parameter dan range nilainya tertentu.

- Kromosom float, adalah kromosom yang disusun dari gen-gen yang bernilai pecahan, termasuk gen yang bernilai bulat. Kromosom ini merupakan model yang jumlah parameternya banyak. Tingkat keberhasilan dari kromosom ini rendah dalam kecepatan.
  - Kromosom string, adalah kromosom yang disusun dari gen-gen yang bernilai string.
  - Kromosom kombinatorial, adalah kromosom yang disusun dari gen-gen yang dinilai berdasarkan urutannya.
- 4) Individu adalah kumpulan gen, bisa dikatakan sama dengan kromosom.
  - 5) Populasi adalah sekumpulan individu yang akan diproses secara bersama-sama dalam satu siklus proses evolusi.
  - 6) Generasi menyatakan satu satuan siklus proses evolusi.
  - 7) Nilai fitness menyatakan seberapa baik nilai dari suatu individu atau solusi yang didapatkan. Nilai yang dijadikan acuan untuk mencapai nilai optimal. Algoritma genetika bertujuan untuk mencari individu yang mempunyai nilai fitness yang paling optimal (bisa maksimum atau minimum, tergantung pada kebutuhan).

### **II.1.2. Mekanisme dan Struktur Umum Algoritma Genetika**

Siklus perkembangbiakan algoritma genetika diawali dengan pembuatan himpunan solusi secara acak (*random*) yang dinamakan populasi, dimana didalamnya terdapat individu-individu yang dinamakan kromosom. Kromosom ini akan secara lambat laun mengalami iterasi perkembangbiakan dalam sebuah generasi. Selama dalam sebuah generasi, kromosom-kromosom ini dievaluasi dengan menggunakan rumus-rumus yang telah ditentukan dalam fungsi fitness. Untuk menciptakan generasi berikutnya dengan generasi yang baru (dinamakan keturunan – *offspring*) dapat dilakukan dengan menggabungkan dua kromosom yang telah didapat sebelumnya dengan menggunakan operator pindah silang (*crossover*) atau dengan memodifikasi sebuah kromosom dengan menggunakan operator mutasi. Sebuah generasi baru sebelum dievaluasi lagi, maka dia melalui proses seleksi berdasarkan fungsi fitnessnya. Dari seleksi ini, kromosom-kromosom yang paling fit (nilai maksimum atau minimum) mempunyai kemungkinan besar untuk terseleksi. Setelah beberapa generasi, algoritma akan mengalami konvergen pada kromosom terbaik, yang merupakan nilai optimum dari permasalahan yang diselesaikan (*Sam'ani, 2012*) [10].

Berikut ini akan dijelaskan ilustrasi struktur algoritma genetika secara umum :



**Gambar 2.1. Ilustrasi Struktur Umum Algoritma Genetika**

Gambar 2.1. menjelaskan bahwa untuk menyelesaikan suatu permasalahan, algoritma genetika diawali dengan menginisialisasikan himpunan solusi yang dibangkitkan secara acak (*random*). Himpunan solusi ini disebut populasi (*Population*). Setiap individu pada populasi disebut kromosom (*Chromosom*), yang menggambarkan sebuah solusi dari suatu masalah yang akan diselesaikan. Sebuah kromosom dapat dinyatakan dalam simbol *string*, misalnya kumpulan *string bit*. Dalam sebuah populasi, setiap kromosom akan dievaluasi dengan menggunakan

alat ukur yang disebut dengan *fitness* (tingkat kesesuaian). Nilai *fitness* ini digunakan untuk mencari dua kromosom (yang memiliki nilai *fitness* yang sesuai) dari sebuah populasi yang akan dijadikan sebagai kromosom induk untuk melakukan regenerasi. Kromosom induk ini akan melakukan regenerasi melalui pindah silang (*crossover*) dan melakukan mutasi (*mutation*) yang akan menghasilkan kromosom baru (*offspring*). Pindah silang (*crossover*) dilakukan dengan cara menggabungkan dua kromosom induk dengan menggunakan operator pindah silang (*crossover*). Sedangkan mutasi hanya berlaku pada sebuah kromosom, dan kromosom ini akan mengalami suatu perubahan (misalnya : 11101100 menjadi 11001100 pada string bit).

Hasil dari pindah silang dan mutasi ini (*offspring*) akan di evaluasi dengan menggunakan alat ukur yang disebut *fitness* (tingkat kesesuaian). Kemudian akan dilihat apakah *offspring* ini merupakan solusi yang optimal atau belum. Jika optimal maka *offspring* inilah jawabannya. Jika tidak, maka *offspring* ini akan diseleksi (*selection*) lagi dengan menggunakan salah satu metode seleksi. *Offspring* yang lulus seleksi akan menjadi populasi yang baru dan akan melakukan regenerasi lagi, sedangkan yang tidak lulus seleksi akan dibuang. Regenerasi akan berhenti jika jumlah iterasi telah terpenuhi dan ditemukannya solusi optimal dari permasalahan yang diselesaikan.

### II.1.3. Komponen Utama Algoritma Genetika

Terdapat delapan komponen utama yang harus dilakukan dalam mengimplementasikan algoritma genetika (Son Kuswadi, 2007) [11], yaitu:

1). Teknik *Encoding/Decoding* Gen dan Individu

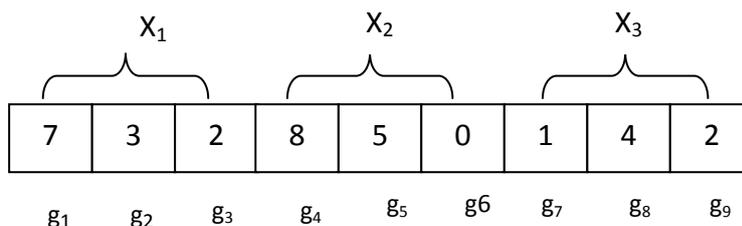
*Encoding* (pengkodean) berguna untuk mengkodekan nilai gen-gen pembentuk individu. Nilai-nilai gen ini diperoleh secara acak. Ada 3 pengkodean yang paling umum digunakan, yaitu :

- a. Pengkodean bilangan real yaitu nilai gen berada dalam interval  $[0,1]$ , sebagai contoh 3 variabel ( $x_1, x_2, x_3$ ) dikodekan kedalam individu yang terdiri dari 3 gen.

$x_1$	$x_2$	$x_3$
0,2431	0,9846	0,5642
$g_1$	$g_2$	$g_3$

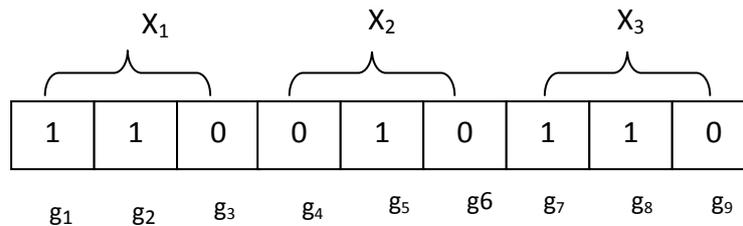
**Gambar 2.2. Contoh Pengkodean Bilangan Real**

- b. Pengkodean Distrik Desimal, yaitu nilai gen berupa bilangan bulat dalam interval  $[0 9]$ , sebagai contoh 3 variabel ( $x_1, x_2, x_3$ ) dikodekan kedalam individu yang terdiri dari 9 gen, tiap-tiap variabel dikodekan kedalam 3 gen.



**Gambar 2.3. Contoh Pengkodean Bilangan Distrik Desimal**

- c. Pengkodean biner yaitu nilai gen berupa bilangan biner 0 atau 1, sebagai contoh 3 variabel ( $X_1$ ,  $X_2$ ,  $X_3$ ) dikodekan ke dalam individu yang terdiri dari 9 gen, tiap-tiap variabel dikodekan kedalam 3 gen.



**Gambar 2.4. Contoh Pengkodean Bilangan Biner**

- 2). Membangkitkan Populasi awal

Sebelum membangkitkan populasi awal, terlebih dahulu kita harus menentukan jumlah individu dalam populasi tersebut. Misalnya jumlah individu tersebut  $N$ . Setelah itu, baru kita membangkitkan populasi awal yang mempunyai  $N$  individu secara acak (*random*). Pencarian solusi dimulai dari satu titik uji tertentu secara acak, titik uji tersebut dianggap sebagai alternative solusi yang disebut sebagai populasi.

- 3). Nilai Fitness

Nilai fitness menyatakan nilai dari fungsi tujuan. Algoritma genetika bertujuan untuk memaksimalkan nilai fitness. Jika yang dicari nilai maksimal, maka nilai fitness adalah nilai dari fungsi itu sendiri. Tetapi jika yang dibutuhkan adalah nilai minimum, maka nilai fitness merupakan invers dari nilai fungsi itu sendiri. Proses invers dapat dilakukan dengan cara berikut.

$$\text{Fitness} = C - f(x) \text{ atau } \text{Fitness} = \frac{C}{f(x) + \epsilon} \quad (2.1)$$

Keterangan :

C adalah konstanta dan  $\epsilon$  adalah bilangan kecil yang ditentukan untuk menghindari agar tidak terjadi pembagian oleh nol dan x adalah individu.

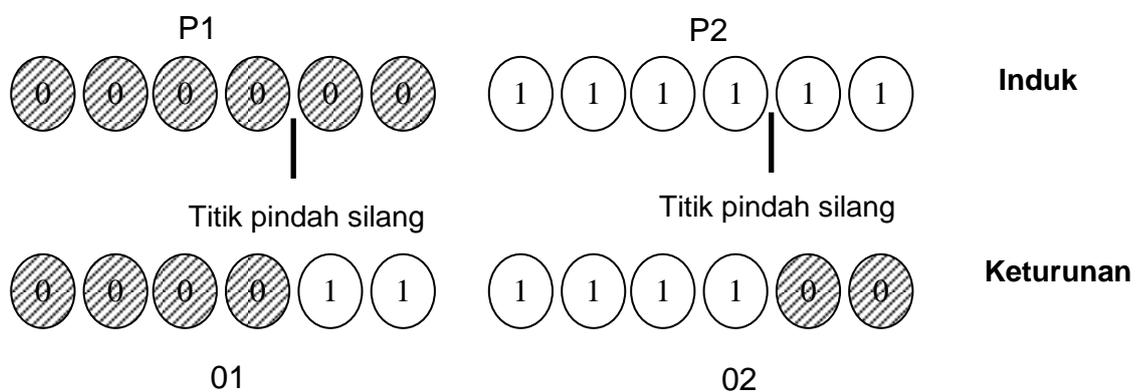
#### 4) Seleksi (*Selection*)

Pada proses evolusi algoritma genetika, keragaman populasi dan tekanan seleksi (*selection pressure*) memegang peran penting. Seleksi merupakan proses yang bertanggung jawab atas pemilihan kromosom dalam proses reproduksi. Metode seleksi yang umumnya digunakan adalah *roulette-wheel*. Metode seleksi dengan mesin roulette ini merupakan metode yang paling sederhana dan sering dikenal dengan nama *stochastic sampling with replacement*. Cara kerja metode ini adalah sebagai berikut:

- a) Hitung total *fitness* semua individu
- b) Hitung probabilitas seleksi masing-masing individu
- c) Dari probabilitas tersebut, dihitung jatah interval masing-masing individu pada angka 0 sampai 1
- d) Bangkitkan bilangan random antara 0 sampai 1
- e) Dari bilangan random yang dihasilkan, tentukan urutan untuk populasi baru hasil proses seleksi.

### 5) Pindah Silang (*Crossover*)

Operator ini bekerja dengan mengambil dua individu dan memotong string kromosom pada posisi terpilih secara acak, untuk memproduksi dua segment *head* dan dua segment *tail*. Sebagai contoh adalah jika kita mengambil induk yang dipresentasikan dengan 5 dimensi vector ( $a_1, b_1, c_1, d_1, e_1$ ) dan ( $a_2, b_2, c_2, d_2, e_2$ ) kemudian dilakukan *crossing* pada posisi ketiga kromosom-kromosomnya sehingga didapat keturunan ( $a_1, b_1, c_2, d_2, e_2$ ) dan ( $a_2, b_2, c_1, d_1, e_1$ ). Berikut ilustrasi dari operasi pindah silang, yaitu :



**Gambar 2.5. Ilustrasi Operasi Pindah Silang Dalam Algoritma Genetika**

### 6). Mutasi (*Mutation*)

Operator mutasi digunakan untuk melakukan modifikasi satu atau lebih nilai gen dalam individu yang sama. Mutasi ini berperan untuk menggantikan gen yang hilang dari populasi akibat seleksi yang memungkinkan munculnya kembali gen yang tidak muncul pada

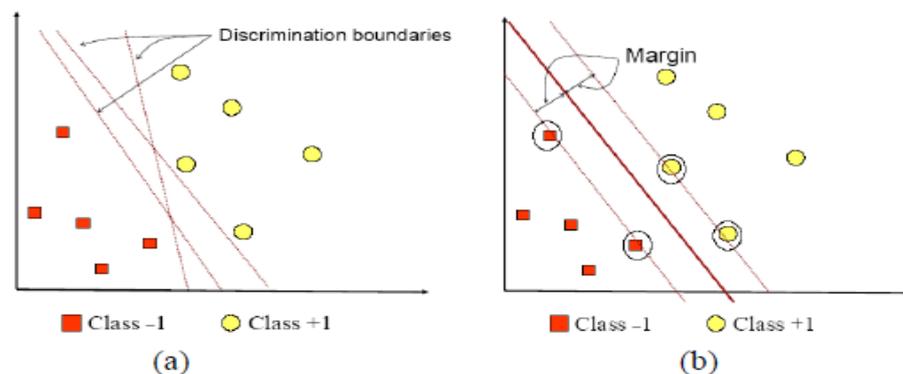
inisialisasi populasi. Metode mutasi yang digunakan adalah mutasi dalam pengkodean nilai. Proses mutasi dalam pengkodean nilai dapat dilakukan dengan berbagai cara, salah satunya yaitu dengan memilih sembarang posisi gen pada kromosom, nilai yang ada tersebut kemudian dirubah dengan suatu nilai tertentu yang diambil secara acak.

## **II.2. Support Vector Machine**

Support Vector Machine (SVM) adalah metode yang digunakan untuk memprediksi, baik dalam kasus klasifikasi maupun regresi. Metode ini diperkenalkan oleh Boser, Guyon, Vapnik pada tahun 1992 di Annual Workshop on Computational Learning Theory. Dalam teknik ini, kita berusaha untuk menemukan fungsi pemisah (*classifier*) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda. *Support Vector Machine (SVM)* dikenal sebagai teknik pembelajaran mesin (*machine learning*) setelah teknik pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Network (NN)*, kedua metode ini telah berhasil digunakan dalam pengenalan pola. Pembelajaran dilakukan dengan menggunakan pasangan data input dan data output berupa sasaran yang diinginkan. Pembelajaran dengan cara ini disebut dengan pembelajaran terarah (*supervised learning*). Dengan pembelajaran terarah ini akan diperoleh fungsi yang menggambarkan bentuk ketergantungan input dan outputnya. Selanjutnya, diharapkan fungsi yang diperoleh mempunyai kemampuan generalisasi yang baik, dalam arti bahwa fungsi tersebut dapat digunakan untuk data input di luar data pembelajaran. Support

Vector Machine telah diterapkan dalam berbagai aplikasi seperti pengolahan citra, pengolahan audio, klasifikasi dokumen web, peramalan (forecasting), dan sebagainya.

Prinsip dasar SVM adalah pengklasifikasi linier, dan selanjutnya dikembangkan agar dapat bekerja pada permasalahan nonlinier. dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi. Perkembangan ini memberikan rangsangan minat penelitian di bidang pengenalan pola untuk investigasi potensi kemampuan SVM secara teoritis maupun dari segi aplikasi. Berikut akan digambarkan proses metode *support vector machine* berusaha menemukan fungsi pemisah (*hyperplane*) terbaik yang memisahkan kedua kelas (Budi Santoso) [12], yaitu :



**Gambar 2.6. SVM Berusaha Menemukan *Hyperplane* Terbaik Yang Memisahkan Kedua Class -1 dan +1**

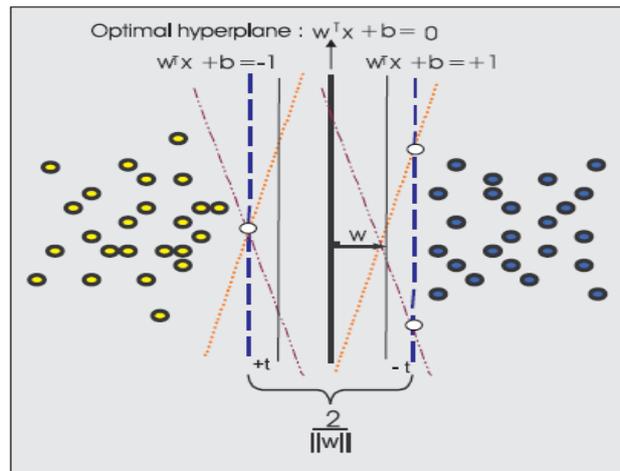
Gambar 2.5. menjelaskan beberapa pola yang merupakan anggota dari dua buah kelas : +1 dan -1. Pola yang tergabung pada class -1 disimbolkan dengan warna merah (kotak), sedangkan pola pada class +1, disimbolkan dengan warna kuning (lingkaran). Problem klasifikasi dapat

diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Kasus klasifikasi yang secara linier bisa dipisahkan, dengan fungsi sebagai berikut :

$$g(x) := \text{sgn}(f(x)) \quad (2.2)$$

$$\text{dengan } f(x) = w^T x + b,$$

dimana  $x, w \in \mathfrak{R}^2$  dan  $b \in \mathfrak{R}$ . Jika ingin menemukan set parameter  $(w,b)$  sehingga  $f(x_i) = \langle w, x \rangle + b = y_i$  untuk semua  $i$ . Teknik ini berusaha berusaha menemukan fungsi pemisah (*hyperplane*) terbaik diantara fungsi yang tidak terbatas jumlahnya untuk memisahkan dua macam obyek. Hyperplane terbaik adalah hyperplane yang terletak ditengah-tengah antara dua set obyek dari dua kelas. Mencari hyperplane terbaik ekuivalen dengan memaksimalkan margin atau jarak antara dua set obyek dari kelas yang berbeda. Jika  $w x_1 + b = +1$  adalah hyperplane-pendukung (supporting hyperplane) dari kelas +1 ( $w x_1 + b = +1$ ) dan  $w x_2 + b = -1$  hyperplane-pendukung dari kelas -1 ( $w x_2 + b = -1$ ), margin antara dua kelas dapat dihitung dengan mencari jarak antara kedua hyperplane-pendukung dari kedua kelas. Gambar berikut akan memperlihatkan bagaimana Support Vector Machine bekerja untuk menemukan suatu fungsi pemisah dengan margin yang maksimal, yaitu :



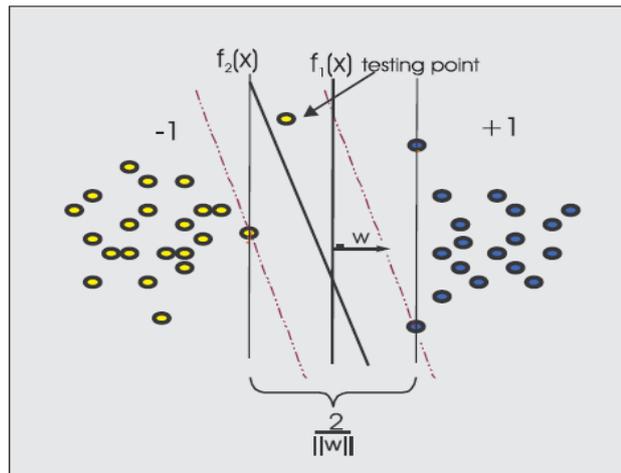
**Gambar 2.7. Mencari Fungsi Pemisah Yang Optimal**

Gambar 2.6. Menjelaskan fungsi  $f_1$  mempunyai margin yang lebih besar dari pada fungsi  $f_2$ . Setelah menemukan dua fungsi ini, sekarang suatu data baru masuk dengan keluaran -1. Pengelompokan data harus dilakukan untuk mengetahui apakah data ini ada dalam kelas -1 atau +1 mengenai fungsi pemisah yang sudah ditemukan.

Secara matematika, formulasi masalah optimisasi SVM untuk kasus klasifikasi linier di dalam *primal space* adalah :

$$\min \frac{1}{2} \|w\|^2 \quad (2.3)$$

$$y_i(w x_i + b) \geq 1, \quad i = 1, \dots, l$$



**Gambar 2-8. Memperbesar Margin Untuk Meningkatkan Probabilitas Suatu Data Secara Benar.**

Dimana  $x_i$  adalah data input,  $y_i$  adalah keluaran dari data  $x_i$ ,  $w$ ,  $b$  adalah parameter-parameter yang kita cari nilainya. Dalam formulasi di atas, ingin fungsi tujuan (obyektif function)  $\min \frac{1}{2} \|w\|^2$  akan diminimalkan atau kuantitas  $\|w\|^2$  atau  $w^T w$  akan dimaksimalkan dengan memperhatikan pembatas  $y_i(w x_i + b) \geq 1$ . Bila output data  $y_i = +1$ , maka pembatas menjadi  $(w x_i + b) \geq 1$ . Sebaliknya bila  $y_i = -1$ , pembatas menjadi  $(w x_i + b) \leq -1$ . Di dalam kasus yang tidak feasible (infeasible) dimana beberapa data mungkin tidak bisa dikelompokkan secara benar, formulasi matematikanya menjadi berikut :

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l t_i \quad (2.4)$$

$$y_i(w x_i + b) + t_i \geq 1$$

$$t_i \geq 0, i = 1, \dots, l$$

dimana  $t_i$  adalah variabel *slack*. Dengan formulasi ini kita ingin memaksimalkan margin antara dua kelas dengan meminimalkan  $\|w\|^2$ .

Dalam formulasi ini kita berusaha meminimalkan kesalahan klasifikasi (misclassification error) yang dinyatakan dengan adanya variabel *slack*  $t_i$ , sementara dalam waktu yang sama kita memaksimalkan margin,  $\frac{1}{\|w\|}$ . Penggunaan variabel *slack*  $t_i$  adalah untuk mengatasi kasus ketidaklayakan (infeasibility) dari pembatas (constraints)  $y_i(w x_i + b) \geq 1$  dengan cara memberi pinalti untuk data yang tidak memenuhi pembatas tersebut. Untuk meminimalkan nilai  $t_i$  ini, kita berikan pinalti dengan menerapkan konstanta ongkos  $C$ . Vektor  $w$  tegak lurus terhadap fungsi pemisah  $w x + b = 0$ . Konstanta  $b$  menentukan lokasi fungsi pemisah relatif terhadap titik asal (origin).

Persamaan (2.4) adalah masalah nonlinear. Ini bisa dilihat dari fungsi tujuan (objective function) yang berbentuk kuadrat. Untuk menyelesaikannya, secara komputasi agak sulit dan perlu waktu lebih panjang. Untuk membuat masalah ini lebih mudah dan efisien untuk diselesaikan, masalah ini bisa kita transformasikan ke dalam *dual space*. Untuk itu, pertama kita ubah persamaan (2.4) menjadi fungsi Lagrangian :

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^l \alpha_i [y_i(w^T x_i + b) - 1] \quad (2.5)$$

dimana variabel *non-negatif*  $\alpha_i$ , dinamakan *Lagrange multiplier*. Solusi dari problem optimisasi dengan pembatas seperti di atas ditentukan dengan mencari *saddle point* dari fungsi Lagrangian  $J(w, b, \alpha)$ . Fungsi ini harus diminimalkan terhadap variabel  $w$  dan  $b$  dan harus dimaksimalkan terhadap variabel  $\alpha$ . Kemudian kita cari turunan pertama dari fungsi  $J(w,$

$b, \alpha$ ) terhadap variabel  $w$  dan  $b$  dan kita samakan dengan 0. Dengan melakukan proses ini, kita akan mendapatkan dua kondisi optimalitas berikut:

1. Kondisi 1 :  $\frac{\partial J(w,b,\alpha)}{\partial w} = 0$

2. Kondisi 2 :  $\frac{\partial J(w,b,\alpha)}{\partial b} = 0$

Penerapan kondisi optimalitas 1 pada fungsi Lagrangian (2.5) akan menghasilkan :

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.6)$$

Penerapan kondisi optimalitas 2 pada fungsi Lagrangian (2.5) akan menghasilkan :

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (2.7)$$

Menurut duality theorem [1]:

1. Jika problem primal mempunyai solusi optimal, maka problem dual juga akan mempunyai solusi optimal yang nilainya sama.
2. Bila  $w_0$  adalah solusi optimal untuk problem primal dan  $\alpha_0$  untuk problem dual, maka syarat *perlu* dan *cukup* bahwa  $w_0$  solusi layak untuk problem primal adalah  $\Phi(w_0) = J(w_0, b_0, \alpha_0) = \min_w J(w, b, \alpha)$ .

Untuk mendapatkan problem dual dari problem dijabarkan persamaan (2.5) sebagai berikut :

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^l \alpha_i y_i w^T x_i - b \sum_{j=1}^l \alpha_j y_j + \sum_{i=1}^l \alpha_i \quad (2.8)$$

Menurut kondisi optimalitas ke dua dalam (2.8), term ketiga sisi sebelah kanan dalam persamaan di atas sama dengan 0. Dengan memakai nilai-nilai  $w$  di (2.6), kita dapatkan :

$$w^T w = \sum_{i=1}^l \alpha_i y_i w^T x_i = \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2.9)$$

maka persamaan (2.8) menjadi :

$$Q(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j x_i^T x_j \quad (2.10)$$

Selanjutnya kita dapatkan formulasi dual dari persamaan (2.10) :

$$\text{Max} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j x_i^T x_j \quad (2.11)$$

$$\sum_{i=1}^l \alpha_i y_i = 0$$

$$0 \leq \alpha_i, i = 1, \dots, l$$

Dengan dot product  $x_i x_j$  sering diganti dengan simbol  $K$ .  $K$  adalah matrik kernel yang dijelaskan dalam bagian 3. Formulasi (2.11) adalah quadratic programming (QP) dengan pembatas (constraint) linier. Melatih SVM ekuivalen dengan menyelesaikan problem *convex optimization*. Karena itu solusi dari SVM adalah unik (dengan asumsi bahwa  $k$  adalah positive definite) dan global optimal. Hal ini berbeda dengan solusi neural networks yang ekuivalen dengan problem *nonconvex optimization* dengan akibat solusi yang ditemukan adalah *local optima*.

$$\text{Ambil } f(x) = \sum_{i=1}^l y_i \alpha_i^* k(x_i, x) + b^*.$$

Fungsi pemisah optimal adalah  $g(x) = \text{sign} (\sum_{i=1}^l y_i \alpha_i^* k(x_i, x)) + b^*$ , dimana  $\alpha_i^*, i = 1, \dots, l$  adalah solusi optimal dari persamaan (10) dan  $b^*$  dipilih sehingga  $y_i f(x_i) = 1$  untuk sembarang  $i$  dengan  $C > \alpha_i^* > 0$ . Data  $x_i$

dimana  $\alpha_i^* > 0$  dinamakan *support vector* dan menyatakan data training yang diperlukan untuk mewakili fungsi keputusan yang optimal. Dalam gambar 3, sebagai contoh, 3 titik berwarna putih menyatakan *support vector*. Untuk mengatasi masalah ketidaklinieran (nonlinearity) yang sering terjadi dalam kasus nyata, kita bisa menerapkan metoda kernel.

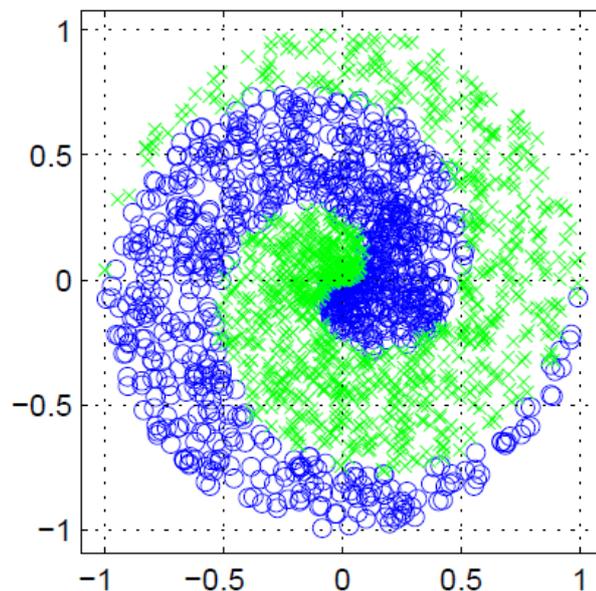
Metoda kernel memberikan pendekatan alternatif dengan cara melakukan mapping data  $x$  dari input space ke *feature space*  $F$  melalui suatu fungsi  $\varphi$  sehingga  $\varphi : x \rightarrow \varphi(x)$ . Karena itu suatu titik  $x$  dalam input space menjadi  $\varphi(x)$  dalam feature space.

Banyak teknik data mining atau machine learning yang dikembangkan dengan asumsi kelinieran. Sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier. Karena itu, bila suatu kasus klasifikasi memperlihatkan ketidaklinieran, algoritma seperti perceptron tidak bisa mengatasinya. Secara umum, kasus-kasus di dunia nyata adalah kasus yang tidak linier.

Dengan metoda kernel suatu data  $x$  di ruang input dipetakan ke ruang feature  $F$  dengan dimensi yang lebih tinggi melalui pemetaan  $\varphi$  sebagai berikut  $\varphi : x \rightarrow \varphi(x)$ . Karena itu data  $x$  di ruang input menjadi  $\varphi(x)$  di ruang feature.

Sering kali fungsi  $\varphi(x)$  tidak tersedia atau tidak bisa dihitung. tetapi *dotproduct* dari dua vektor dapat dihitung baik di dalam ruang input maupun di ruang *feature*. Dengan kata lain, sementara  $\varphi(x)$  mungkin tidak diketahui, dot product  $\langle \varphi(x_1), \varphi(x_2) \rangle$  masih bisa dihitung di ruang

feature. Untuk bisa memakai metoda kernel, pembatas (constraint) perlu diekspresikan dalam bentuk dot product dari vektor data  $x_i$ . Sebagai konsekuensi, pembatas yang menjelaskan permasalahan dalam klasifikasi harus diformulasikan kembali sehingga menjadi bentuk dot product. Gambar 2-8 menjelaskan suatu contoh pemetaan feature dari ruang dua dimensi ke ruang feature dua dimensi.



**Gambar 2-9. Data Spiral Yang Menggambarkan Ketidaklinieran**

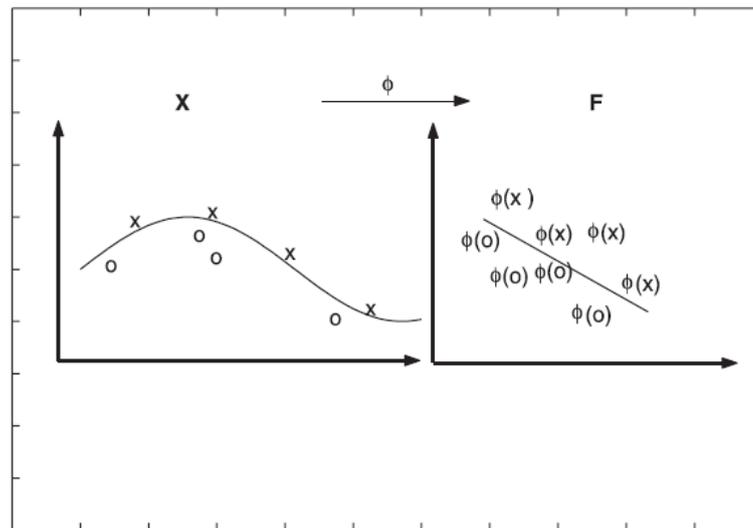
Dalam ruang input, data tidak bisa dipisahkan secara linier, tetapi kita bisa memisahkan di ruang feature. Karena itu dengan memetakan data ke ruang feature Karena itu dengan memetakan data ke ruang feature menjadikan tugas klasifikasi menjadi lebih mudah.

Fungsi kernel yang biasanya dipakai dalam literatur SVM :

- linier :  $x^T x$ ,

- Polynomial :  $(x^T x_i + 1)^p$ ,
- Radial basis function (RBF) :  $\exp(-\frac{1}{2\sigma^2} \|x - x_i\|^2)$
- Tangent hyperbolic (sigmoid) :  $\tanh(\beta x^T x_i + \beta_1)$ , dimana  $\beta, \beta_1 \in R$

?



**Gambar 2-10. Suatu Kernel Map Mengubah Problem Yang Tidak Linier Menjadi Linier Dalam Space Baru**

Fungsi kernel mana yang harus digunakan untuk substitusi dot product di feature space sangat bergantung pada data. Biasanya metoda cross-validation digunakan untuk pemilihan fungsi kernel ini. Pemilihan fungsi kernel yang tepat adalah hal yang sangat penting. Karena fungsi kernel ini akan menentukan ruang feature di mana fungsi klasifier akan dicari. Sepanjang fungsi kernelnya legitimate, SVM akan beroperasi secara benar meskipun kita tidak tahu seperti apa map yang digunakan. Fungsi kernel yang legitimate diberikan oleh Teori Mercer dimana fungsi itu harus memenuhi syarat: kontinu dan positive definite. Lebih mudah

menemukan fungsi kernel daripada mencari map  $\varphi$  seperti apa yang tepat untuk melakukan mapping dari ruang input ke ruang feature. Pada penerapan metoda kernel, kita tidak perlu tahu map apa yang digunakan untuk satu per satu data, tetapi lebih penting mengetahui bahwa dot produk dua titik di feaure space bisa digantikan oleh fungsi kernel.

Peramalan (Forecasting) suatu nilai dapat diperkirakan dengan menggunakan regresi. Teknik regresi yang sering digunakan dalam berbagai bidang, contohnya regresi linear sederhana dan regresi linear multi variabel. Support Vector Machine ini berusaha mencari pemisah antar-kelas (*hyperplane*) yang maksimum. Kelebihan teknik ini dibandingkan dengan teknik klasifikasi lain adalah meminimalkan *error* pada *training-set* dan tingkat generalisasi dari SVM tidak dipengaruhi oleh dimensi dari input vektor (Nugroho *et al.* 2003)

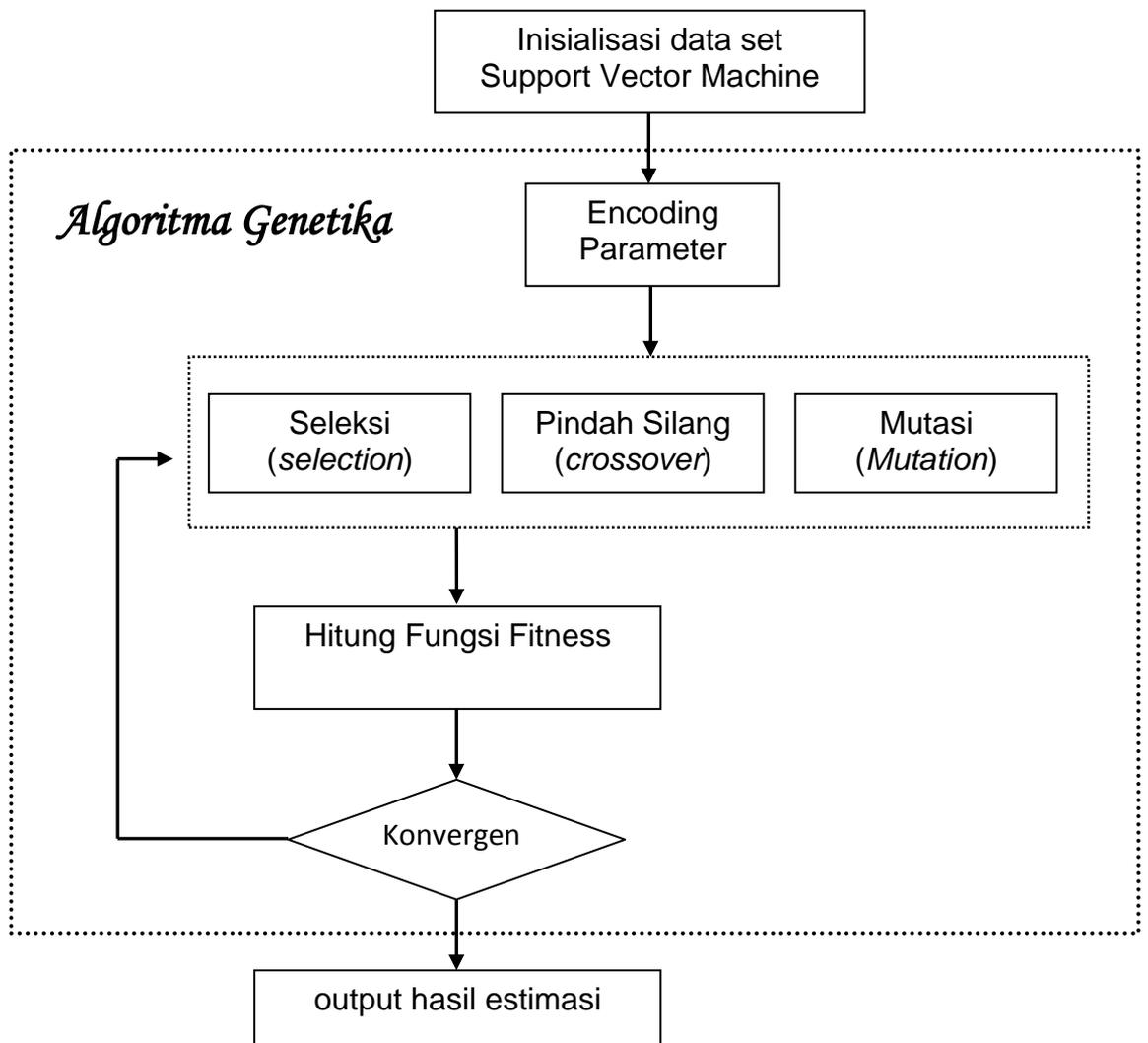
### **II.3. Kombinasi Algoritma Genetika dengan Support Vector Machine**

Peramalan time series masih menerima perhatian yang serius dari komunitas penelitian. Penerapan teknik kecerdasan komputasi untuk memecahkan masalah prediksi atau estimasi dengan menganalisis kumpulan data (data set) masa lampau (*history*) . Analisis ini dimungkinkan oleh karena kemajuan teknik komputasi dan teknologi informasi. Salah satu teknik komputasi yaitu mesin pembelajaran (*learning machine*) untuk memprediksi data time-series adalah Support Vector Machine (SVM). Aplikasi ini melakukan analisis data berdasarkan pada

indikator yang berasal dari data yang relevan. Dalam rangka meningkatkan akurasi prediksi atau estimasi, terlebih dahulu diketahui indikatornya. Dalam kondisi tertentu misalnya meramalkan populasi pohon dalam suatu hutan, lebih dari 100 indikator telah dikembangkan untuk memahami perilaku terhadap kondisi hutan. Dalam kasus seperti ini untuk mengoptimalkan algoritma yang sudah diterapkan harus dikaji untuk mengidentifikasi indikator yang diperlukan.

Algoritma genetika adalah algoritma yang paling umum digunakan dan memiliki kemampuan untuk menangani fungsi tujuan yang memiliki banyak kendala. Algoritma genetika digunakan untuk memilih masukan terbaik untuk metode Support Vector Machine dalam peramalan berdasarkan kumpulan data (*data set*). Optimasi diperlukan dalam metode Algoritma Genetika kombinasi Support Vector Machine adalah mewujudkan atau menghapus data input yang tidak diperlukan untuk estimasi atau menghindari kerentanan data yang biasa terjadi pada model dengan terlalu banyak parameter.

Berikut ini akan digambarkan struktur umum, kombinasi Algoritma Genetika dengan Support Vector Machine (SVM), yaitu :



**Gambar 2-11. Struktur Umum Kombinasi Algoritma Genetika Dengan Support Vector Machine**

Dari gambar 2.11. dijelaskan proses kerja dari kombinasi Algoritma Genetika dengan Support Vector Machine (SVM), sebagai berikut :

1. Inisialisasi Data Set

Inisialisasi data set dilakukan untuk mengelompokkan data-data berdasarkan pemicu.

## 2. Encoding parameters

Dalam penelitian ini, parameter pemicu pengurang luas lahan hutan dan penambah luas lahan hutan diwakili oleh kromosom yang terdiri bilangan desimal.

## 3. Proses yang dilakukan Algoritma Genetika

### a) Seleksi Chromosom

Algoritma genetika melakukan proses seleksi terhadap data yang telah dikodekan. Pada umumnya proses penyeleksian dilakukan dengan menggunakan mesin *roulette-wheel*. Cara kerja metode ini adalah menghitung total fitness semua individu (masing-masing data pemicu penambah luas areal hutan dan pemicu pengurang luas areal hutan), menghitung probabilitas seleksi masing-masing individu, dan membangkitkan bilangan acak (*random*) antara 0 sampai 1.

### b) Pindah silang (*Crossover*)

Setelah proses seleksi selesai maka selanjutnya algoritma melanjutkan proses pindah silang (*crossover*). Salah satu cara yang digunakan adalah one-cut-point, yaitu memilih secara acak satu posisi dalam kromosom induk kemudian saling menukar gen.

Berikut pseudo-code proses crossover, adalah :

Begin

$k \leftarrow 0$ ;

While ( $k < \text{populasi}$ ) do

$R[k] \leftarrow \text{random}(0-1)$ ;

If ( $R[k] < pc$ ) then

    Select Chromosome[k] as parent;

End;

#### c) Mutasi

Proses ini dilakukan dengan cara mengganti satu gen yang terpilih secara acak dengan satu nilai baru yang didapat secara acak. Pertama dihitung dahulu panjang total gen yang ada dalam satu populasi. Rumus Total gen = (jumlah gen dalam kromosom) \* jumlah populasi.

#### 4. Hitung fungsi fitness

Setelah proses genetika dilakukan dan mendapatkan nilai pada generasi baru, maka proses selanjutnya dengan melakukan perhitungan fungsi fitness baru

Rumus fungsi fitness = (kebakaran hutan + perambahan + illegal logging) - reboisasi.

#### 5. Perbandingan Nilai

Proses ini dilakukan untuk membandingkan nilai fungsi fitness dengan nilai dari generasi baru. Jika nilai fungsi fitness sama dengan nilai dari generasi baru, maka konvergensi akan terjadi. Nilai terakhir

yang diperoleh dijadikan nilai hasil estimasi luas areal hutan. Dan jika belum konvergen atau sama nilainya maka akan dilakukan pengulangan pada nomor 2, sampai nilai konvergen tercapai.

#### **II.4. Basis Data (*Database*)**

basis data adalah kumpulan file-file yang mempunyai kaitan antara satu file dengan file yang lain sehingga membentuk satu bangunan data sebagai informasi untuk suatu perusahaan atau instansi dalam batasan tertentu.

Susunan basis data dibagi menjadi beberapa tingkatan data, yaitu sebagai berikut :

a. Sistem basis data

Sistem basis data merupakan kumpulan basis data dalam suatu sistem yang mungkin tidak berhubungan satu sama lain, namun secara umum mempunyai hubungan sistem. Secara sederhana sistem basis data tersusun dari banyak file.

b. Basis data

Basis data merupakan kumpulan bermacam-macam record yang mempunyai hubungan antar record, agregat data, dan field terhadap satu objek tertentu.

c. File

File merupakan kumpulan record sejenis secara relasi.

d. Record

Record merupakan kumpulan dari field atau agregat data yang saling berhubungan dengan suatu objek tertentu.

e. Field

Field merupakan unit terkecil yang disebut data, item, atau elemen data.

f. Byte

Byte merupakan bagian terkecil yang dialamatkan pada memori. Byte adalah kumpulan bit yang secara konvensional terdiri dari 8 bit. Satu byte dipakai untuk mengkodekan satu buah karakter dalam memori.

g. Bit

Sistem angka biner terdiri dari dua nilai saja yaitu 0 dan 1. Sistem angka biner merupakan dasar komunikasi antara manusia dan mesin.

Semua data yang tersimpan dalam basis data dengan sebuah perangkat lunak sistem manajemen basis MySQL. MySQL tersedia sebagai perangkat lunak gratis di bawah lisensi *general public licence* (GPL).

## II.5. MySQL

MySQL adalah sebuah sistem manajemen database relasi (relational data-base management system) yang bersifat “terbuka” (*open source*). Terbuka maksudnya MySQL boleh didownload oleh siapa saja, baik versi kode program aslinya (source code program) maupun versi binernya (excuteble program).

Keunggulan MySQL antara lain sebagai berikut:

a. Portability

MySQL dapat berjalan stabil pada berbagai sistem operasi.

b. Open source

MySQL dapat didistribusikan secara open source (gratis), dibawah lisensi General Public License (GPL) sehingga dapat digunakan secara Cuma-Cuma.

c. Multiuser

MySQL dapat digunakan oleh beberapa user dalam waktu yang bersamaan tanpa mengalami masalah. Hal ini memungkinkan sebuah database server MySQL dapat diakses client secara bersamaan.

d. Performance tuning

MySQL memiliki kecepatan yang dalam menangani query sederhana, dengan kata lain dapat memproses lebih banyak SQL persatuan waktu.

e. Colum types

MySQL memiliki tipe kolom yang sangat kompleks, seperti signed integer, float, double, char, varchar, text, blob, data, time, timestamp, year, set serta enum.

f. Command dan functions

MySQL memiliki operator dan fungsi secara penuh yang mendukung perintah SELECT dan WHERE dalam query.

g. Security

MySQL memiliki beberapa lapisan sekuritas seperti level subnetmask, nama host, dan ijin akses user dengan sistem perijinan yang mendetail serta password terenkripsi.

h. Scalability dan limits

MySQL mampu menangani database dalam skala besar, dengan jumlah records lebih dari 50 juta dan 60 ribu tabel serta 5 miliar baris. Selain itu batas indeks yang dapat ditampung mencapai 32 indeks pada tiap tabelnya.

## **II.6. Unified Modeling Language (UML)**

*Unified Modeling Language* (UML) adalah alat atau bahasa pemodelan yang sering digunakan untuk melakukan konstruksi dan dokumentasi dari aplikasi atau perangkat lunak (software) berbasis objek yang akan dibangun, dan dilakukan sebelum pengkodean (*coding*). Tujuan melakukan pemodelan dengan unified modeling Language (UML) dalam merancang atau mengembangkan piranti lunak adalah untuk memenuhi semua kebutuhan pengguna dengan lengkap dan tepat,

dengan mempertimbangkan faktor skalabilitas (*scalability*), ketangguhan (*robustness*) dan keamanan (*security*).

Tujuan UML menurut Booch, Rumbaugh dan Jacobson :

- Memberikan model yang siap pakai, bahasa pemodelan visual yang ekspresif untuk mengembangkan dan saling menukar model dengan mudah dan dimengerti secara umum.
- Memberikan bahasa pemodelan yang bebas dari berbagai bahasa pemrograman dan proses rekayasa.
- Menyatukan praktek-praktek terbaik yang terdapat dalam pemodelan.

Dengan menggunakan UML kita dapat membuat model untuk semua jenis aplikasi piranti lunak, dimana aplikasi tersebut dapat berjalan pada piranti keras, sistem operasi dan jaringan apapun, serta ditulis dalam bahasa pemrograman apapun.

## II.7. Road Map Penelitian

Beberapa peneliti terdahulu yang telah melakukan penelitian tentang prediksi keadaan atau kejadian dimasa yang akan datang, antara lain :

1. **“Application of GA-SVM Time Series Prediction in Tax Forecasting”**, oleh Sheng Lu, Zhong-jian Cai, Chongqing Technology and Business University Chongqing, China, 2009. Penelitian ini menggunakan aplikasi Algoritma Genetika (GA) dengan Support Vector Machine (SVM) untuk meramalkan pajak keuntungan kotor dari ekonomi pasar. Model konvensional linear ekonomi makroskopik

sangat sulit untuk menahan fenomena non-linier di sistem ekonomi, sehingga kesalahan peramalan pajak akan meningkat. Dukungan support vector machine (SVM) telah berhasil digunakan untuk memecahkan masalah regresi non-linier dan sampel kecil. Berdasarkan aritmatika regresi SVM, dengan algoritma genetika (GA-SVM) diusulkan untuk memperkirakan pajak, dimana algoritma genetika (GA) digunakan untuk menentukan parameter pelatihan dukungan vektor mesin. Hasil percobaan menunjukkan bahwa model yang diusulkan GA-SVM dapat mencapai akurasi besar di bawah keadaan data pelatihan kecil [12].

Penelitian tersebut di atas menguatkan metode yang digunakan pada penelitian ini, dimana algoritma GA juga digunakan untuk menentukan parameter luas lahan yang akan di klasifikasikan oleh algoritma SVM, sehingga melalui penelitian ini dapat dilihat luas lahan yang dipengaruhi oleh setiap parameter pemicu yang mengurangi luas lahan dan pemicu yang menambah luas lahan.

2. **“Prediction of Urban Water Demand Based on GA-SVM”**, oleh Xiaogang Chen, Huaiyin Institute of Technology, Huaian, Chine, 2010. Penelitian ini memprediksi kebutuhan pasokan air perkotaan dengan dukungan support vector machine (SVM) yang kemudian dioptimalkan oleh algoritma genetika (GA). Algoritma GA digunakan untuk menentukan parameter pelatihan SVM. Hasil eksperimen menunjukkan bahwa model yang diusulkan GA-SVM tidak hanya

membutuhkan data kecil untuk pelatihan, tetapi juga dapat mencapai akurasi besar pada data pelatihan yang berukuran besar. Pada penelitian ini GA-SVM diterapkan untuk meramalkan kebutuhan pasokan air perkotaan. Pada penelitian ini GA digunakan untuk memilih parameter untuk pelatihan pada SVM [13].

Model penggunaan GA-SVM pada penelitian di atas digunakan juga pada penelitian ini, dimana data-data pemicu lahan menjadi parameter SVM sehingga hasil klasifikasi SVM menjadi kromosom untuk proses optimasi menggunakan algoritma GA.

3. **“A new support vector machine- genetic algorithm (SVM-GA) based method for stock market forecasting”**, oleh Vahid Khatibi, Elham Khatibi dan Abdolreza Rasouli, Islamic Azad University, Bardsir, Iran, 2011. Penelitian ini melakukan peramalan dalam bidang pasar modal. Dukungan mesin vektor (SVM) digunakan untuk menafsirkan data keuangan. Tapi SVM tunggal tidak dapat mencapai hasil yang akurat. Selanjutnya, dalam penelitian ini SVM dikombinasikan algoritma genetika (GA). Kombinasi algoritma ini tujuan untuk peramalan pasar saham. Algoritma GA berguna untuk memilih indikator input sebagai fitur input untuk memprediksi arah masa depan harga saham. Hasil penelitian menunjukkan bahwa konsep korelasi GA membantu dalam meningkatkan kinerja sistem SVM secara signifikan [14]. Pada penelitian di atas parameter dibidang keuangan diklasifikasikan menggunakan SVM dan prediksi

menjadi optimal menggunakan GA. Penerapan GA-SVM pada penelitian ini mempunyai model yang sama dengan penelitian di atas, tetapi berbeda pada objek penerapan model tersebut dimana pada penelitian ini data-data pemicu yang mempengaruhi luas lahan di klasifikasikan oleh SVM dan kemudian diprediksi oleh GA.

4. **“GA-SVM Based Framework for Time Series Forecasting”**, oleh Thi Nguyen, Lee Gordon-Brown, Peter Wheeler, Jim Peterson, Monash University, Australia, 2009. Penelitian ini menggunakan pendekatan normal untuk prediksi time-series dimana algoritma GA terintegrasi. SVM dimasukkan ke dalam model GA yang memiliki peningkatan yang luar biasa dalam masalah perkiraan dan akurasi karena dapat mengetahui masukan yang diperlukan dan menghapus masukan yang tidak perlu. Umumnya, metode integrasi antara GA dan SVM seperti studi ini juga dapat diterapkan untuk teknik pembelajaran mesin lainnya. Peningkatan efisiensi yang handal diperoleh dalam model GA-SVM karena biaya komputasi tidak intensif dan jauh lebih rendah dari pada teknik pembelajaran mesin lainnya [15]. Pada penelitian di atas, menguatkan metode yang digunakan pada penelitian ini, sehingga model GA-SVM merupakan sebuah model yang terintegrasi untuk peramalan luas areal hutan TNK.
5. **“Penghitungan laju luas area hutan berbasis algoritma segmentasi warna local”**, oleh Dyah Pratiwi, dkk., Universitas Gunadarma, 2013.

Penelitian ini bertujuan menghitung laju perubahan (penambahan atau pengurangan) luas area hutan. Penelitian ini menggunakan algoritma segmentasi penginderaan jarak jauh atau *remote sensing*. Tahapan kegiatan yang dilakukan meliputi pencuplikan piksel warna referensi hutan, transformasi warna dari RGB ke dalam ruang warna HSV, ekstraksi dan segmentasi berbasis warna melalui disimilaritas jarak warna pada jendela partisi, proses binerisasi dan penelusuran tepi objek melalui rantai Freeman yang dilanjutkan penghitungan jumlah objek dan luas area objek keseluruhan. Hasil penghitungan luas objek pada suatu periode dibandingkan dengan penghitungan luas objek pada periode lain. Metode ini menjadi acuan dan pemikiran untuk pengambilan data secara langsung di lokasi penelitian [16].

7. **"Forest Area Estimation Using Sample Survey and Landsat MSS and TM Data"**, oleh F. Deppe, Federal University of Rio Grande do Sul, American, 1998. Estimasi area hutan pada penelitian ini menggunakan metode regresi dengan sumber data citra satelit (Landsat). Estimasi area hutan dengan metode estimator regresi dengan cara mengoreksi rata-rata estimasi variabel Y (tutupan lahan tertentu) sebagai fungsi dari hasil yang diperoleh dari variabel X (proporsi piksel yang dihasilkan dari klasifikasi citra). Metode Estimator regresi digunakan untuk menghasilkan perkiraan objektif dari klasifikasi data survey tanah melalui penggunaan citra satelit [17].

Penelitian di atas menggunakan data yang berasal dari citra satelit dengan metode estimator regresi, sementara pada penelitian ini sumber data berasal dari data sekunder luas areal hutan TNK dengan estimasi luas area hutan menggunakan algoritma GA-SVM.

8. **“Forecasting Areas Vulnerable to Forest Conversion in the Tam Dao National Park Region”**, Oleh Dang Khoi dan Yuji Murayama, University of Tsukuba, 2010. Penelitian ini memprediksi daerah rentan terhadap perubahan hutan dengan menggunakan metode multi-layer perceptron neural jaringan (MLPNN) dengan model rantai Markov (MLPNN-M). Model MLPNN-M memprediksi peningkatan tekanan pada hutan primer yang tersisa dalam taman serta pada hutan sekunder di daerah sekitarnya. Hasil Penelitian dapat digunakan untuk memprioritaskan lokasi bagi upaya pengelolaan hutan konservasi keanekaragaman hayati masa depan. Metode ini menjadi konsep pemikiran dalam penelitian yang diusulkan untuk mengetahui pola perubahan hutan. Menurut penelitian tersebut faktor reboisasi perlu diperhitungkan [18].

9. **“A Bayesian approach to multi-source forest area estimation”**, oleh Andrew O. Finley, Sudipto Banerjee, Ronald E. McRoberts, USA, 2007.

Penelitian ini menyajikan model hirarkis Bayesian yang memungkinkan untuk analisis dan plot data hutan. Hasil prediksi menunjukkan model yang diusulkan memberikan akurasi klasifikasi

yang tinggi 88%. Pendekatan yang dilakukan menjadi input dalam penelitian yang diusulkan untuk mengetahui teknik plot data [19].

10. **“Estimation of Basal Area in West Oak Forests of Iran Using Remote Sensing Imagery”**, Loghman Ghahramany, Pariz Fatehi, Hedayat Ghazanfari, University of Kurdistan, Iran, 2010. Tujuan penelitian ini adalah untuk mengevaluasi kemampuan citra satelit untuk meramalkan areal hutan di Northern Zagros, Iran [20].
11. **“AUTOMATIC ESTIMATION OF FOREST INVENTORY PARAMETERS BASED ON LIDAR, MULTI-SPECTRAL AND FOGIS DATA”**, oleh Oliver Diederhagen, Barbara Koch, Holgen Weinacker, University Freiburg, Germany, 2003. Penelitian ini menggunakan metode teknologi penginderaan jauh, dengan tujuan adalah untuk menentukan variabel dasar untuk inventarisasi hutan seperti jenis pohon, tinggi pohon. Parameter ini digunakan untuk memperkirakan umur pohon, diameter batang, dan volume kayu. Metode ini menjadi konsep pemikiran dalam penelitian yang diusulkan untuk mengetahui parameter-parameter yang digunakan dalam estimasi areal hutan [21].
12. **“Importance of Vegetation Type in Forest Cover Estimation”**, oleh Anuj Karpatne, dkk, University of Minnesota, 2010. Penelitian ini menggunakan metode K-Means pada data time series dengan pengamatan disetiap lokasi selama periode 5 tahun (2004-2005),

untuk masing-masing kluster yang diperoleh sesuai dengan jenis vegetasi yang berbeda [22].

13. **“Using A Sampling Method For Estimation for Forest Canopy cover”**, oleh Farshad Keivan Behjou, Mahbobeh Foshat, University of Mohaghegh Ardabili, Iran, 2013. Penelitian ini menggunakan metode sampling berdasarkan teori probabilitas untuk mengestimasi tutupan lahan daerah hutan [23].
14. **“Estimation of Forest Canopy Cover : a Comparison of field measurement Techniques”**, oleh Lauri Korhonen, Kari T. Khorhonen, Miina Rautianen and Pauline Stenberg, 2006. Penelitian ini mengestimasi tutupan lahan hutan dengan cara mengambil data dari camera (foto digital). Metode ini memberikan input dalam penelitian ini tentang bagaimana mengambil data melalui camera [24].