



**OPTIMASI K-NEAREST NEIGHBOUR MENGGUNAKAN  
GENETIC ALGORITHM PADA SISTEM KLASIFIKASI  
WHITE SPOT DISEASE**

**PROPOSAL SKRIPSI**

diajukan guna memenuhi salah satu syarat  
untuk melaksanakan seminar proposal

**Oleh:**

**Zainal Hasan**

**NIM 172410101006**

**PROGRAM STUDI SISTEM INFROMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS JEMBER**

**2020**

## A. Judul

Optimasi *K-Nearest Neighbour* Menggunakan *Genetic Algorithm* pada Sistem Klasifikasi *White Spot Disease*

## B. Latar Belakang

Negara Indonesia merupakan salah satu negara dengan kawasan kepulauan terbesar di dunia. Oleh karena itu banyak sekali dilakukan budidaya perairan di Indonesia. Udang merupakan salah satu komoditas unggulan yang memiliki nilai ekspor terbesar sekitar 21% hasil perikanan di perdagangan dunia. (Samah, 2018)

Udang *Vannamei* berasal dari perairan Amerika dan Hawaii yang sukses dikembangkan di berbagai negara di Asia. Udang *Vannamei* merupakan salah satu spesies unggul pengganti udang Windu. Jenis udang ini mulai dibudidayakan di Indonesia sejak tahun 2002. Pada awalnya pemilihan udang *Vannamei* dibandingkan udang Windu dikarenakan udang *Vannamei* cukup tahan dari serangan penyakit. Namun dalam perkembangannya, udang *Vannamei* mulai terserang penyakit salah satunya adalah *White Spot Disease* (WSD) atau lebih dikenal penyakit bercak putih yang disebabkan oleh *White Spot Syndrome Virus* (WSSV). (Tancung & Kordi, 2007)

Karena kemunculan penyakit ini, maka dibutuhkan solusi untuk menganalisis penyakit ini secara cepat dan mudah untuk mempercepat pengambilan tindakan pencegahan. Untuk membantu proses klasifikasi ini, petani udang memerlukan sarana yang dapat membantu mengakomodir kebutuhan tersebut. Dengan adanya kemajuan teknologi di masa kini, proses ini dapat diselesaikan dengan memanfaatkan *artificial intelligence* dan *data mining*. Sistem klasifikasi penyakit merupakan salah satu jenis sistem pakar yang dapat melakukan proses klasifikasi berdasarkan sekelompok data dan menghasilkan informasi baru.

Dengan adanya *historical data* terkait penyakit ini, maka akan lebih mudah proses analisis dilakukan menggunakan *data mining*. *Data mining* merupakan salah satu teknik yang bisa digunakan untuk melakukan *improve human judgement* dari sekumpulan data sebagai salah satu proses analisa otomatis untuk mengetahui pola atau kecenderungan dari sekelompok data yang kemudian dapat diolah menjadi sebuah informasi baru. Pada *data mining* terdapat banyak sekali fungsi yang dapat digunakan salah satunya yaitu fungsi klasifikasi.

Salah satu algoritma untuk menyelesaikan klasifikasi dengan mudah adalah algoritma *K-Nearest Neighbour*. Algoritma jenis ini salah satu algoritma yang sederhana dan mudah dipelajari, tahan dengan data yang memiliki *noise* dan efektif jika datanya cukup besar. Namun, algoritma ini juga memiliki kekurangan yaitu proses komputasi yang kompleks, keterbatasan memori, dapat tertipu dengan atribut yang tidak relevan dan juga parameter K yang bias. (Siti Mutrofin et al., 2014) Untuk mencegah nilai parameter K bias, maka dapat dilakukan optimasi penentuan parameter K. Terdapat banyak sekali algoritma yang dapat digunakan untuk optimasi seperti *Bat Algorithm*, *Bird Swarm Algorithm*, *Particle Swarm Optimization*, *Multi Swarm Optimization*, *Genetic Algorithm*, dan masih banyak lagi lainnya.

Dari bermacam-macam jenis algoritma optimasi diatas, *Genetic Algorithm* atau lebih dikenal algoritma Genetika merupakan salah satu teknik optimasi yang melakukan proses optimasi secara acak berdasarkan prinsip evolusi dan mampu melakukan penelusuran yang kompleks. (Suguna & Thanushkodi, 2010) Dengan dilakukannya optimasi pada sistem klasifikasi ini, diharapkan dapat meningkatkan akurasi dari hasil klasifikasi sehingga dapat dimanfaatkan dengan lebih baik lagi.

### **C. Rumusan Masalah**

Berdasarkan uraian pada latar belakang di atas dapat diidentifikasi beberapa masalah penelitian sebagai berikut:

1. Bagaimana cara melakukan optimasi pada sistem klasifikasi *White Spot Disease* pada udang *Vannamei* menggunakan *Genetic Algorithm*?
2. Bagaimana hasil optimasi pada proses klasifikasi *White Spot Disease* pada udang *Vannamei* menggunakan *Genetic Algorithm*?

### **D. Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah dipaparkan sebelumnya, maka tujuan penelitian ini sebagai berikut:

1. Mengetahui alur implementasi optimasi pada proses klasifikasi *White Spot Disease* pada udang *Vannamei* menggunakan *Genetic Algorithm*
2. Mengetahui hasil optimasi pada proses klasifikasi *White Spot Disease* pada udang *Vannamei* menggunakan *Genetic Algorithm*

## E. Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah dipaparkan sebelumnya, maka diharapkan penelitian ini dapat memberikan manfaat sebagai berikut:

1. Bagi Peneliti

Dengan dilakukannya penelitian ini diharapkan peneliti dapat menerapkan ilmu pengetahuan yang telah didapatkan di perkuliahan utamanya dalam bidang Agroindustri sesuai dengan visi Fakultas Ilmu Komputer, Universitas Jember.

2. Bagi Akademisi

Hasil penelitian ini diharapkan dapat menambah referensi bagi peneliti lain di masa yang akan datang.

## F. Batasan Masalah

Agar tidak terjadi penyimpangan dalam proses penelitian, maka ditetapkan Batasan masalah dalam penelitian ini. Adapun batasan masalah dalam penelitian ini, antara lain:

1. Penelitian ini berfokus pada proses optimasi algoritma KNN menggunakan algoritma Genetika
2. Data yang akan dijadikan objek penelitian merupakan *dataset* dengan sumber terbuka yang digunakan pada penelitian “*A sequential assessment of WSD risk factors of shrimp farming in Bangladesh: Looking for a sustainable farming system*” (Hasan et al., 2020)

## G. Tinjauan Pustaka

Bagian ini menyajikan penelitian-penelitian terdahulu yang relevan dengan optimasi klasifikasi baik yang menggunakan metode genetika maupun yang lain serta objek yang sama, seperti yang disajikan berikut ini.

### G.1. Penelitian Terdahulu

Penelitian terdahulu yang terkait dengan proses optimasi klasifikasi pada algoritma *K-Nearest Neighbour* yang menggunakan *Genetic Algorithm* maupun algoritma optimasi lainnya dapat dilihat pada tabel berikut:

|    |                  |   |
|----|------------------|---|
| 1. | Nama             | N. Suguna, Dr. K. Thanushkodi   |
|    | Tahun            | 2010  |
|    | Judul            | <i>An Improved k-Nearest Neighbor Classification Using Genetic Algorithm</i>  |
|    | Analisis         | Algoritma KNN merupakan salah satu algoritma yang sederhana dan memiliki kompatibilitas tinggi. Namun, algoritma KNN juga memiliki kelemahan sehingga diperlukan proses optimasi pada setiap permasalahan.  |
|    | Hasil Penelitian | Secara mendasar, proses optimasi ini dilakukan dengan cara menghitung kemiripan antar atribut sebanyak K tetangga kemudian menghitung akurasi. Hasilnya, kompleksitas dari proses penghitungan menjadi lebih kecil dan proses menjadi berjalan lebih cepat. Selain itu, proses pembobotan pada setiap sampel tidak lagi diperlukan dengan hasil akurasi yang semakin meningkat. |
| 2. | Nama             | Siti Mutrofin, Abidatul Izzah, Arrie Kurniawardhani, Mukhamad Masrur  |
|    | Tahun            | 2014  |
|    | Judul            | Optimasi Teknik Klasifikasi <i>Modified K Nearest Neighbor</i> Menggunakan Algoritma Genetika   |
|    | Analisis         | Penelitian ini dilakukan untuk melakukan perbaikan pada algoritma MKNN dengan cara melakukan optimasi nilai K menggunakan algoritma Genetika  |

|  |                  |  |
|--|------------------|--|
|  | Hasil Penelitian | <p>Algoritma KNN, MKNN dan GMKNN memiliki kinerja yang sama baiknya, dalam melakukan klasifikasi data Iris dengan hasil akurasi 100%.</p> <p>Keunggulan dari algoritma GMKNN adalah dapat menentukan nilai K optimal pada MKNN dengan otomatis, tanpa harus mencoba satu persatu dalam menentukan nilai K.</p> |
|--|------------------|--|

## G.2. *White Spot Disease*

*White Spot Disease* atau lebih dikenal penyakit bercak putih merupakan salah satu penyakit yang memiliki inang cukup luas. Penyakit ini disebabkan oleh *White Spot Syndrome Virus*. Udang yang terinfeksi virus ini dapat mengalami kematian hingga 100%. Oleh karenanya, jika ditemukan adanya tanda-tanda munculnya penyakit ini, maka harus segera dilakukan panen sebelum menimbulkan kerugian yang lebih besar.

## G.3. *Data Mining*

*Data mining* adalah proses menemukan pola atau informasi menarik dari sekelompok data menggunakan teknik, algoritma atau metode tertentu. Menurut (Larose, 2005), *data mining* dapat menyelesaikan 6 *task* yaitu:

### 1. *Description*

Fungsi *description* berguna untuk membantu mendeskripsikan suatu pola dan tren dari suatu data. Fungsi ini dapat menghasilkan deskripsi kualitas tinggi dengan cara melakukan analisis data eksplorasi.

### 2. *Estimation*

Fungsi *estimation* digunakan untuk memperkirakan nilai numerik dari suatu atribut berdasarkan sekumpulan variabel numerik.

### 3. *Prediction*

Fungsi *prediction* tidak jauh berbeda dengan fungsi *classification* dan fungsi *estimation* kecuali dalam sisi hasilnya. Hasil dari fungsi *prediction* ini bisa bernilai benar ataupun salah di masa yang akan datang.

4. *Classification*

Fungsi *classification* tidak jauh berbeda dengan fungsi *estimation* kecuali pada variabel yang berupa variabel kategorisasi.

5. *Clustering*

Fungsi *clustering* bertujuan dalam mengelompokkan suatu data berdasarkan data yang memiliki kemiripan. Sebuah *cluster* adalah sekelompok data yang mirip satu sama lain.

6. *Association*

Fungsi *association* bertujuan untuk mencari asosiasi dari beberapa atribut. Hal ini bertujuan untuk melakukan analisis keterkaitan antar atribut.

#### **G.4. *K-Nearest Neighbour Algorithm***

*K-Nearest Neighbors Algorithm* atau algoritma KNN merupakan salah satu algoritma sederhana untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Algoritma KNN lebih banyak digunakan karena algoritma ini mudah dipahami dan proses penghitungannya sederhana.

Algoritma ini akan memprediksi data *test* dengan data *training* sebanyak K yang mana K merupakan jumlah tetangga terdekat dengan data *test*, kemudian memasukkannya ke dalam kategori yang memiliki probabilitas paling besar. (Suguna & Thanushkodi, 2010) Syarat nilai K adalah tidak boleh lebih besar dari jumlah data *training*. Selain itu, nilai K harus bernilai ganjil dan lebih dari satu.

Jauh atau dekatnya jarak titik dengan titik tetangganya dapat dihitung dengan menggunakan *Euclidean Distance*. Penghitungan *Euclidean Distance* dapat dinyatakan dengan formula matematika berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Salah satu kekurangan algoritma ini adalah penentuan nilai K yang tidak tepat dapat mengurangi akurasi hasil klasifikasi. Oleh karena itu terdapat banyak sekali metode yang dapat membantu algoritma KNN agar semakin akurat.

### **G.5. Genetic Algorithm**

*Genetic Algorithm* atau algoritma Genetika merupakan salah satu algoritma yang mencontoh proses evolusi alami, sehingga hanya individu yang unggul yang akan bertahan hidup dan yang lemah akan punah. Untuk mengukur tingkat keunggulan dari suatu individu digunakan fungsi *fitness*. Fungsi ini berguna untuk mengetahui kecocokan atau kemiripan antar individu berdasarkan tiap kromosomnya. Algoritma Genetika ini sering digunakan untuk melakukan pencarian data ataupun untuk menyelesaikan permasalahan optimasi.

Siklus dari algoritma Genetika dimulai dengan mempersiapkan populasi, kemudian dilakukan penghitungan *fitness* antar individu. Proses selanjutnya adalah melakukan seleksi untuk mencari individu yang terbaik. Kemudian, dilakukan *crossover* terhadap setiap individu yang terbaik. Setelah itu dilanjutkan dengan proses mutasi dari setiap individu unggul tersebut sehingga dihasilkan populasi baru. Siklus ini akan terus diulang untuk menghasilkan populasi baru yang lebih baik dari sebelumnya dan dilakukan hingga generasi ke-n.

## **H. Metodologi Penelitian**

### **H.1. Jenis Penelitian**

Penelitian yang akan dilakukan ini merupakan penelitian eksperimen. Penelitian eksperimen merupakan sebuah metode penelitian yang bertujuan untuk mengetahui pengaruh suatu perlakuan tertentu terhadap yang lain dalam kondisi yang terkendali. Metode penelitian eksperimen ini merupakan metode inti dari penelitian yang mengambil pendekatan kuantitatif karena terdapat proses mengontrol, kegiatan manipulasi dan observasi.



## H.2. Tahapan Penelitian

Tahapan penelitian pada penelitian ini digambarkan dalam diagram berikut:



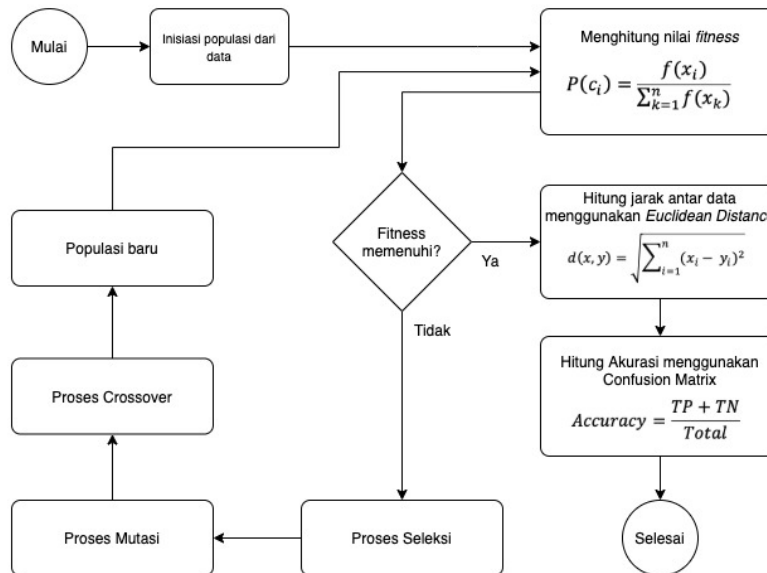
## H.3. Data dan Parameter

Parameter yang digunakan pada penelitian ini adalah jenis pakan, kepadatan daya tampung, sumber air, frekuensi pertukaran air, rentang waktu pembuangan lumpur, penggunaan bahan kimia untuk persiapan kolam, penggunaan bahan kimia untuk pengolahan air, salinitas air, keasaman air, jenis tanah kolam, penggunaan lahan kolam sebelumnya, prevalensi penyakit bercak putih pada periode sebelumnya, dan metode budidaya. Hasil klasifikasi berupa resiko munculnya penyakit bercak putih.

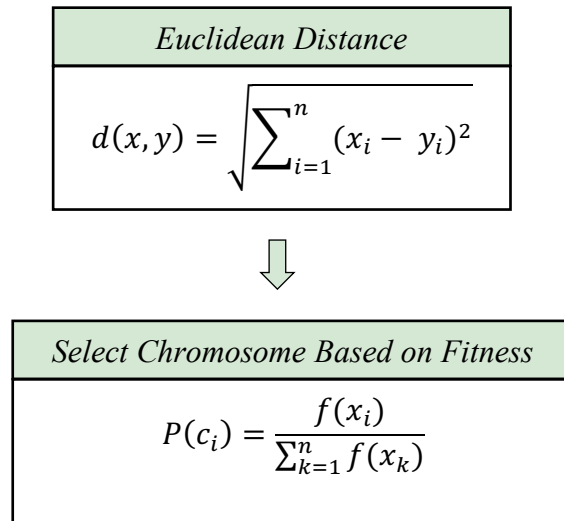
#### H.4. Rancangan Sistem

Sistem yang akan dibangun berupa web yang dapat mengklasifikasikan resiko munculnya penyakit bercak putih pada udang *Vannamei*. Sistem ini akan mendapat *input* berupa parameter yang sudah ditentukan pada sub-bab sebelumnya. Nilai *input*-an berdasarkan kondisi terkini di tambak yang akan dilakukan klasifikasi. Kemudian, sistem akan memberi informasi resiko munculnya *White Spot Disease* pada tambak tersebut. Kemudian, pada sistem klasifikasi ini akan dilakukan optimasi menggunakan algoritma Genetika. Nantinya, hasil klasifikasi akan dibandingkan antara sebelum dilakukan optimasi dengan setelah dilakukan optimasi.

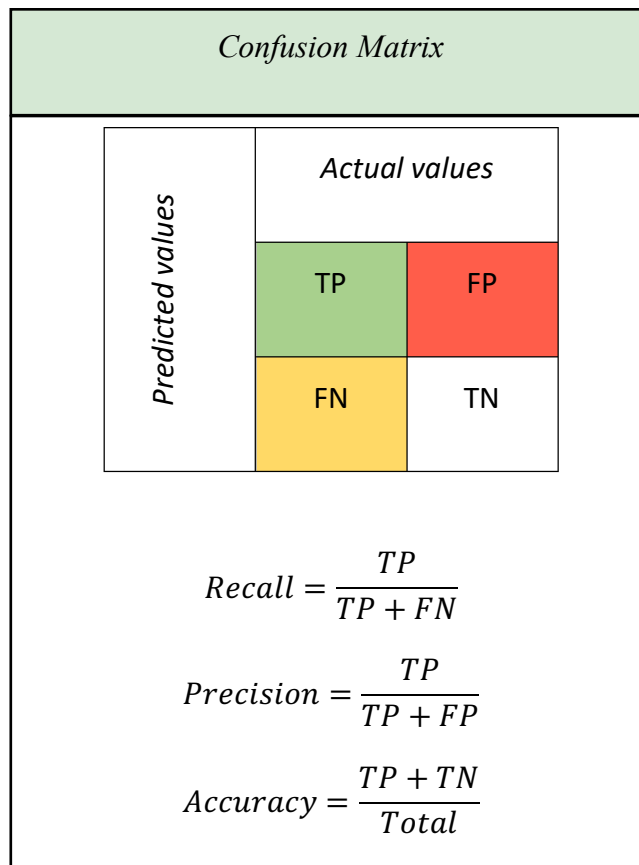
Gambaran dari alur sistem dapat digambarkan sebagai berikut:



Gambaran dari alur penghitungan dalam sistem dapat digambarkan sebagai berikut:



Untuk mengetahui tingkat keberhasilan optimasi akan diukur berdasarkan nilai akurasi sebelum dan sesudah optimasi. Nilai akurasi didapatkan dengan menggunakan *Confusion Matrix*. Model matematika dari *confusion matrix* dapat digambarkan sebagai berikut:



## H.5. Luaran yang Diharapkan

Dari penelitian ini, diharapkan luaran sebagai berikut:

1. Skripsi
2. Sistem klasifikasi *White Spot Disease* pada udang *Vannamei* menggunakan algoritma KNN

## I. Jadwal Kegiatan

Rencana jadwal kegiatan disajikan pada tabel berikut ini

| No. | Kegiatan  | Bulan |     |     |     |     |
|-----|---|-------|-----|-----|-----|-----|
|     |   | Nov   | Des | Jan | Feb | Mar |
| 1   | Studi Literatur                                     |       |     |     |     |     |
| 2   | Penentuan Populasi dan Sampel                       |       |     |     |     |     |
| 3   | Penentuan Jenis dan Sumber Data                     |       |     |     |     |     |
| 4   | Penentuan Variabel dan Skala Pengukuran Klasifikasi |       |     |     |     |     |
| 5   | Perancangan Model Klasifikasi dan Optimasi          |       |     |     |     |     |
|     | Implementasi Model Klasifikasi dan Optimasi         |       |     |     |     |     |
|     | Pengujian Model Klasifikasi dan Hasil Optimasi      |       |     |     |     |     |

## J. Daftar Pustaka

- Hasan, N. A., Haque, M. M., Hinchliffe, S. J., & Guilder, J. (2020). A sequential assessment of WSD risk factors of shrimp farming in Bangladesh: Looking for a sustainable farming system. *Aquaculture*, 526, 735348. <https://doi.org/10.1016/j.aquaculture.2020.735348>
- Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. In

*Discovering Knowledge in Data: An Introduction to Data Mining.* Wiley.  
<https://doi.org/10.1002/0471687545>

Samah, A. E. (2018). PROSPEK PERDAGANGAN UDANG INDONESIA. *Buletin Ilmiah Litbang Perdagangan*, 1(2 SE-Articles). <https://doi.org/10.30908/bilp.v1i2.300>

Siti Mutrofin, Izzah, A., Kurniawardhani, A., & Masrur, M. (2014). Optimasi teknik klasifikasi modified k nearest neighbor menggunakan algoritma genetika. *Gamma*, 10(September), 130–134. <http://ejournal.umm.ac.id/index.php/gamma>

Suguna, N., & Thanushkodi, K. (2010). An Improved k-Nearest Neighbor Classification Using Genetic Algorithm. *International Journal of Computer Science Issues*, 7(4), 18–21.

Tancung, A. B., & Kordi, M. G. (2007). *Pengelolaan Kualitas Air dalam Budi Daya Perairan*. Rineka Cipta. <https://books.google.co.id/books?id=M92wtQEACAAJ>