

**SKRIPSI**

**SISTEM PREDIKSI TINGGI GELOMBANG LAUT DI SELAT  
MAKASSAR MENGGUNAKAN DATA METEOROLOGI DAN  
OSEANOGRAFI**

**Disusun dan diajukan oleh:**

**MAGHFIRAH TENRI SUMPALA ZANI  
D121181520**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2023**

**LEMBAR PENGESAHAN SKRIPSI**  
**SISTEM PREDIKSI TINGGI GELOMBANG LAUT DI SELAT**  
**MAKASSAR MENGGUNAKAN DATA METEOROLOGI DAN**  
**OSEANOGRAFI**

**Disusun dan diajukan oleh**

**MAGHFIRAH TENRI SUMPALA ZANI**

**D121181520**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 6 Februari 2023 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

Prof. Dr. Ir. Indrabayu ST, MT, M. Bus. Sys., IPM, ASEAN.  
Nip. 19750716 200212 1 004

Pembimbing Pendamping,

Elly Warni ST., MT  
Nip. 19820216 200812 2 001

Ketua Program Studi,



Prof. Dr. Ir. Indrabayu ST, MT, M. Bus. Sys., IPM, ASEAN  
Nip. 19750716 200212 1 004

## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama : Maghfirah Tenri Sumpala Zani

NIM : D121181520

Program Studi : Teknik Informatika

Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Sistem Prediksi Tinggi Gelombang Laut di Selat Makassar Menggunakan Data Meteorologi dan Oseanografi

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

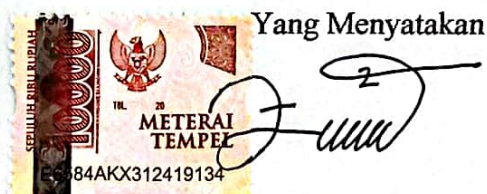
Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 17 Februari 2023

Yang Menyatakan



Maghfirah Tenri Sumpala Zani

## ABSTRAK

**MAGHFIRAH TENRI SUMPALA ZANI.** *Sistem Prediksi Tinggi Gelombang Laut di Selat Makassar Menggunakan Data Meteorologi dan Oseanografi* (dibimbing oleh Prof. Dr. Ir. Indrabayu ST, MT, M.Bus.Sys., IPM, ASEAN. Eng dan Elly Warni ST., MT)

Selat Makassar merupakan Alur Laut Kepulauan Indonesia (ALKI) yang memiliki tingkat lalu lintas laut yang tinggi. Namun, seringkali terjadi insiden kecelakaan laut di Selat Makassar yang diakibatkan oleh tingginya gelombang laut. Faktor terjadinya gelombang laut yaitu akibat cuaca diantaranya angin dan hujan. Selain itu, factor oseanografi juga secara tidak langsung mempengaruhi gelombang laut seperti perubahan suhu permukaan laut akan mengakibatkan perubahan cuaca kemudian mengakibatkan gelombang laut.

Penelitian ini bertujuan untuk menemukan hubungan atau pola item variabel kecepatan angin, arah angin, suhu permukaan laut, curah hujan terhadap tinggi gelombang laut. Berdasarkan pola antara item digunakan untuk menyaring fitur yang dianggap relevan untuk digunakan dalam proses prediksi. Selain itu, pola atau aturan asosiasi yang terbentuk dapat digunakan untuk mendukung hasil prediksi. Penelitian ini juga bertujuan untuk membangun model sistem prediksi tinggi gelombang laut di Selat Makassar.

Pencarian pola atau hubungan antar variabel menggunakan metode *association rule* dengan Algoritma *FP-Growth* dengan penetapan *minimum support*=0.003, *minimum confidence* = 0.98 dan nilai *lift* > 1. Adapun untuk pembangunan model prediksi menggunakan algoritma *Support Vector Regression*.

Hasil asosiasi menunjukkan bahwa parameter yang paling berpengaruh terhadap tinggi gelombang laut di masing-masing bagian di Selat Makassar yaitu parameter angin diantaranya kecepatan angin dan arah angin. Hasil prediksi Pada skenario 1 memiliki rata rata *error* pada *data training* yaitu 0.0623 dan pada *data testing* yaitu sebesar 0.0635. Sedangkan pada skenario 2 memiliki rata rata *error* pada *data training* yaitu 0.0628 dan pada *data testing* yaitu sebesar 0.0635. Hal ini membuktikan bahwa pada model yang dihasilkan oleh skenario 1 lebih unggul dibandingkan skenario 2

Kata Kunci: Gelombang Laut, Asosiasi, Prediksi, *FP-Growth*, *Support Vector Regression*

## ABSTRACT

**MAGHFIRAH TENRI SUMPALA ZANI.** Sea Wave Height Prediction System in Makassar Strait Using Meteorological and Oceanographic Data (supervised by Prof. Dr. Ir. Indrabayu ST, MT, M.Bus.Sys., IPM, ASEAN. Eng and Elly Warni ST, MT)

Makassar Strait is an Indonesian Archipelago Sea Route (ALKI) that has a high level of sea traffic. However, there are often incidents of sea accidents in the Makassar Strait caused by high sea waves. The factors for the occurrence of sea waves are due to weather including wind and rain. In addition, oceanographic factors also indirectly affect sea waves such as changes in sea surface temperature will cause weather changes and then cause sea waves.

This study aims to find the relationship or pattern of variable items of wind speed, wind direction, sea surface temperature, rainfall to sea wave height. Based on the pattern between items, it is used to filter out features that are considered relevant for use in the prediction process. In addition, the patterns or association rules formed can be used to support the prediction results. This research also aims to build a model of the sea wave height prediction system in the Makassar Strait.

The search for patterns or relationships between variables uses the association rule method with the FP-Growth Algorithm with the determination of minimum support = 0.003, minimum confidence = 0.98 and lift value > 1. As for the construction of a prediction model using the Support Vector Regression algorithm.

The association results show that the parameters that have the most influence on sea wave height in each part of the Makassar Strait are wind parameters including wind speed and wind direction. The prediction results in scenario 1 have an average error in the training data of 0.0623 and in the testing data of 0.0635. Meanwhile, scenario 2 has an average error in the training data of 0.0628 and in the testing data of 0.0635. This proves that the model produced by scenario 1 is superior to scenario 2.

Keywords: Sea Wave, Association, Prediction, FP-Growth, Support Vector Regression

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR.....	vi
DAFTAR TABEL.....	vii
DAFTAR SINGKATAN DAN ARTI SIMBOL.....	ix
DAFTAR LAMPIRAN.....	xi
KATA PENGANTAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian/Perancangan.....	3
1.4 Manfaat Penelitian/Perancangan.....	3
1.5 Ruang Lingkup/Asumsi perancangan.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Gelombang Laut.....	4
2.2 Angin.....	6
2.3 Suhu Permukaan Laut.....	8
2.4 Curah Hujan.....	10
2.5 <i>Knowledge Discovery in Database (KDD)</i> .....	10
2.6 <i>Association Rule</i> .....	14
2.7 <i>Algoritma FP-Growth</i> .....	16
2.8 <i>Machine Learning</i> .....	22
2.9 Peramalan ( <i>Forecasting</i> ).....	23
2.10 <i>Support Vector Regression (SVR)</i> .....	23
2.11 <i>Root Mean Square Error (RMSE)</i> .....	26
BAB 3 METODE PENELITIAN/PERANCANGAN.....	28
3.1 Tahapan Penelitian.....	28
3.2 Waktu dan Lokasi Penelitian.....	29
3.3 Instrumen Penelitian.....	29
3.4 Teknik Pengambilan Data.....	30
3.5 Perancangan Sistem.....	31
3.6 Asosiasi menggunakan <i>Algoritma FP-Growth</i> .....	42
3.7 Prediksi menggunakan <i>Support Vector Regression</i> .....	44
3.8 Evaluasi Sistem.....	45
BAB 4. HASIL DAN PEMBAHASAN.....	47
4.1 Penerapan <i>Algoritma FP-Growth</i> .....	47
4.2 Penerapan <i>Algoritma Support Vector Regression</i> .....	63
4.3 Pembahasan.....	80
BAB 5. KESIMPULAN DAN SARAN.....	88
5.1 Kesimpulan.....	88
5.2 Saran.....	88
DAFTAR PUSTAKA.....	90

## DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Gelombang .....	4
Gambar 2. 2 Proses Knowledge Discovery.....	10
Gambar 2. 3 Hasil pembentukan FP-tree TID 1 (Samuel, 2008).....	19
Gambar 2. 4 Hasil pembentukan FP-tree TID 2 (Samuel, 2008).....	19
Gambar 2. 5 Hasil pembentukan FP-tree TID 3 (Samuel, 2008).....	20
Gambar 2. 6 Hasil pembentukan FP-tree TID 10 (Samuel, 2008).....	20
Gambar 2. 7 SVR linier satu dimensi.....	23
Gambar 2. 8 Insensitive zone (a) original input space, dan (b) feature space.....	25
Gambar 3. 1 Tahapan penelitian .....	28
Gambar 3. 2 Titik Kordinat Pengambilan Data .....	30
Gambar 3. 3 Rancangan Sistem .....	31
Gambar 3. 4 Hasil deteksi <i>missing value</i> dataset Selat Makassar bagian selatan .....	37
Gambar 3. 5 Hasil deteksi <i>missing value</i> dataset Selat Makassar bagian tengah.....	37
Gambar 3. 6 Hasil deteksi <i>missing value</i> dataset Selat Makassar bagian utara ....	37
Gambar 3. 7 Flowchart Algoritma FP-Growth .....	42
Gambar 3. 8 Flowchart prediksi.....	44
Gambar 4. 1 Pembangunan FP-tree.....	52
Gambar 4. 2 Contoh Conditional Pattern Base .....	53
Gambar 4. 3 Visualisasi perbandingan data aktual dan prediksi skenario 1 Selat Makassar bagian selatan.....	76
Gambar 4. 4 Visualisasi perbandingan data aktual dan prediksi skenario 1 selat makassar bagian tengah .....	77
Gambar 4. 5 Visualisasi perbandingan data aktual dan prediksi skenario 1 selat makassar bagian utara .....	77
Gambar 4. 6 Visualisasi perbandingan data aktual dan prediksi skenario 2 selat makassar bagian selatan.....	77
Gambar 4. 7 Visualisasi perbandingan data aktual dan prediksi skenario 2 selat makassar bagian tengah .....	78
Gambar 4. 8 Visualisasi perbandingan data aktual dan prediksi selat makassar bagian utara .....	78

## DAFTAR TABEL

Tabel 2. 1 Kategori tinggi gelombang.....	5
Tabel 2. 2 Distribusi kecepatan angin .....	7
Tabel 2. 3 Kategori arah angin.....	7
Tabel 2. 4 Kategori Suhu Permukaan Laut .....	9
Tabel 2. 5 Kategori Curah Hujan .....	10
Tabel 2. 6 Tabel data transaksi mentah.....	17
Tabel 2. 7 Frekuensi kemunculan tiap karakter .....	18
Tabel 2. 8 Tabel data transaksi.....	18
Tabel 3. 1 Sampel dataset dari Badan Meteorologi dan Klimatologi (BMKG)....	32
Tabel 3. 2 Sampel dataset dari Badan Riset dan Inovasi (BRIN) .....	33
Tabel 3. 3 Sampel dataset dari <i>Remote Sensing Systems</i> (REMSS) .....	33
Tabel 3. 4 Sampel hasil seleksi dataset dari BMKG.....	34
Tabel 3. 5 Sampel hasil seleksi dataset dari BRIN .....	34
Tabel 3. 6 Sampel hasil seleksi dataset dari REMSS .....	34
Tabel 3. 7 Sampel hasil integrasi bagian selatan Selat Makassar .....	35
Tabel 3. 8 Sampel hasil integrasi bagian tengah Selat Makassar.....	36
Tabel 3. 9 Sampel hasil integrasi bagian utara Selat Makassar .....	36
Tabel 3. 10 Hasil handling missing value bagian selatan Selat Makassar .....	38
Tabel 3. 11 Hasil handling missing value bagian tengah Selat Makassar .....	38
Tabel 3. 12 Hasil handling missing value bagian utara Selat Makassar .....	38
Tabel 3. 13 Sampel hasil transformasi data Selat Makassar bagian selatan .....	39
Tabel 3. 14 Sampel hasil transformasi data Selat Makassar bagian tengah.....	39
Tabel 3. 15 Sampel hasil transformasi data Selat Makassar bagian utara .....	40
Tabel 3. 16 Sampel hasil <i>binning</i> dataset Selat Makassar bagian selatan.....	40
Tabel 3. 17 Sampel hasil binning dataset Selat Makassar bagian tengah .....	41
Tabel 3. 18 Sampel hasil binning dataset Selat Makassar bagian utara.....	41
Tabel 3. 19 Parameter Tuning dengan metode Grid Search .....	45
Tabel 4. 1 Sampel dataset setelah dilakukan preprocessing data.....	47
Tabel 4. 2 Support tiap item dataset Selat Makassar bagian selatan.....	48
Tabel 4. 3 Support tiap item dataset Selat Makassar bagian tengah .....	48
Tabel 4. 4 Support tiap item dataset Selat Makassar bagian utara.....	49
Tabel 4. 5 Item yang memenuhi <i>minimum support</i> dataset Selat Makassar bagian selatan .....	49
Tabel 4. 6 Tabel item yang memenuhi <i>minimum support</i> dataset Selat Makassar bagian tengah.....	50
Tabel 4. 7 Tabel item yang memenuhi <i>minimum support</i> dataset Selat Makassar bagian utara .....	50
Tabel 4. 8 Tabel data transaksi Selat Makassar bagian selatan.....	51
Tabel 4. 9 Tabel data transaksi Selat Makassar bagian tengah .....	51
Tabel 4. 10 Tabel data transaksi Selat Makassar bagian utara.....	51
Tabel 4. 11 Frequent itemset Selat Makassar bagian selatan.....	54
Tabel 4. 12 Frequent itemset Selat Makassar bagian tengah .....	54
Tabel 4. 13 Frequent itemset Selat Makassar bagian utara .....	54
Tabel 4. 14 Aturan asosiasi gelombang tenang.....	55
Tabel 4. 15 Aturan asosiasi gelombang rendah .....	56



Tabel 4. 16 Aturan asosiasi gelombang sedang .....	57
Tabel 4. 17 Aturan asosiasi gelombang tinggi .....	57
Tabel 4. 18 Aturan asosiasi gelombang tenang.....	58
Tabel 4. 19 Aturan asosiasi gelombang rendah .....	59
Tabel 4. 20 Aturan asosiasi gelombang sedang .....	59
Tabel 4. 21 Aturan asosiasi gelombang tinggi.....	60
Tabel 4. 22 Aturan asosiasi kategori gelombang tenang.....	61
Tabel 4. 23 Aturan asosiasi kategori gelombang rendah .....	61
Tabel 4. 24 Aturan asosiasi kategori gelombang sedang .....	62
Tabel 4. 25 Aturan asosiasi kategori gelombang tinggi.....	62
Tabel 4. 26 Sampel dataset untuk prediksi.....	64
Tabel 4. 27 Sampel data hasil normalisasi skenario 1 .....	65
Tabel 4. 28 Sampel data hasil normalisasi skenario 2 .....	65
Tabel 4. 29 Pembagian data training dan data testing.....	66
Tabel 4. 30 Tabel hasil tuning parameter skenario 1 .....	66
Tabel 4. 31 Tabel hasil tuning parameter skenario 2 .....	67
Tabel 4. 32 Waktu tuning parameter.....	67
Tabel 4. 33 Hasil pencarian variasi pembagian data terbaik.....	68
Tabel 4. 34 Sampel hasil prediksi skenario 1 Selat Makassar bagian selatan.....	69
Tabel 4. 35 Sampel hasil prediksi skenario 1 Selat Makassar bagian tengah .....	69
Tabel 4. 36 Sampel hasil prediksi skenario 1 Selat Makassar bagian utara.....	70
Tabel 4. 37 Hasil prediksi skenario 2 Selat Makassar bagian selatan.....	71
Tabel 4. 38 Hasil prediksi skenario 2 Selat Makassar bagian tengah .....	72
Tabel 4. 39 Hasil prediksi skenario 2 Selat Makassar bagian utara.....	72
Tabel 4. 40 Hasil denormalisasi skenario 1 Selat Makassar bagian selatan .....	73
Tabel 4. 41 Hasil denormalisasi skenario 1 Selat Makassar bagian tengah.....	73
Tabel 4. 42 Hasil denormalisasi skenario 1 Selat Makassar bagian utara .....	74
Tabel 4. 43 Hasil denormalisasi skenario 2 Selat Makassar bagian selatan .....	75
Tabel 4. 44 Hasil denormalisasi skenario 2 Selat Makassar bagian tengah.....	75
Tabel 4. 45 Hasil denormalisasi skenario 2 Selat Makassar bagian utara .....	76
Tabel 4. 46 Waktu prediksi .....	79
Tabel 4. 47 Perhitungan evaluasi model menggunakan RMSE.....	79
Tabel 4. 48 Perbandingan RMSE algoritma SVR dengan Regresi Linear .....	86

## DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
ALKI	Alur Laut Kepulauan Indonesia
ABK	Anak Buah Kapal
AIMP	<i>Artificial Intelligence and Multimedia Processing</i>
B	Barat
BBD	Barat Barat Daya
BBL	Barat Barat Laut
BD	Barat Daya
BL	Barat Laut
BMKG	Badan Meteorologi dan Klimatologi
BRIN	Badan Riset dan Inovasi
C	<i>Complexity</i>
CC	<i>Coefficient Correlation</i>
CSV	<i>Comma Separated Values</i>
FP	<i>Frequent Pattern</i>
HDD	<i>Hard Disk Drive</i>
IHK	Indeks Harga Konsumen
KDD	<i>Knowledge Discovery In Databases</i>
L	Liter
RAM	<i>Random Access Memory</i>
RBF	<i>Radial Basis Function</i>
REMSS	<i>Remote Sensing System</i>
RMSE	<i>Root Mean Square Error</i>
S	Selatan
SBD	Selatan Barat Daya
SSD	<i>Solid State Drive</i>
STG	Selatan Tenggara
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
T	Timur

TB	Terabyte
TG	Tenggara
TL	Timur Laut
TTG	Timur Tenggara
TTL	Timur Timur Laut
U	Utara
UBL	Utara Barat Laut
UTL	Utara Timur Laut
$\alpha$	alpha
$\gamma$	gamma
$\xi$	<i>slack</i>
$\epsilon$	epsilon

## DAFTAR LAMPIRAN

Lampiran 1 Dataset dan Hasil .....	94
Lampiran 2 <i>Source Code</i> .....	95

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan rahmat dan karunia-Nya, sehingga dapat menyelesaikan tugas akhir dengan judul “Sistem Prediksi Tinggi Gelombang Laut di Selat Makassar Menggunakan Data Meteorologi dan Oseanografi” sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 di Departemen Teknik Informatika, Fakultas Teknik Universitas Hasanuddin.

Gagasan yang melatarbelakangi tajuk permasalahan ini timbul dari hasil pengamatan penulis terhadap berita kecelakaan laut di Selat Makassar yang dialami oleh para nelayan pada saat berlayar. Kecelakaan laut yang terjadi di Selat Makassar mayoritas diakibatkan oleh tingginya gelombang laut. Oleh karena itu, penulis bermaksud membuat model prediksi gelombang laut di Selat Makassar untuk memberikan informasi kepada nelayan terkait tinggi gelombang laut demi menghindari kecelakaan laut di Selat Makassar.

Penulis menyadari bahwa penyusunan dan penulisan tugas akhir ini tidak dapat terselesaikan dengan baik tanpa adanya bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih banyak kepada :

1. Keluarga penulis, Bapak Zainal Arifin dan Ibu Hasriani selaku kedua orang tua penulis. Nanda Mutiara Zani selaku kakak kandung penulis yang selalu memberikan dukungan, doa, dan semangat yang tiada hentinya, serta selalu sabar dalam mendidik penulis
2. Bapak Prof. Dr. Ir. Indrabayu ST, MT, M.Bus.Sys., IPM, ASEAN. Eng, selaku pembimbing I dan Ibu Elly Warni, S.T., M.T selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran dan perhatian yang luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
3. Segenap staf dan dosen Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir.
4. Bapak Mahagnyana yang telah senantiasa membantu penulis selama

pengambilan data penelitian.

5. Segenap keluarga AIMP Research Group Universitas Hasanuddin yang telah memberikan begitu banyak bantuan selama penelitian dan diskusi terkait penyusunan tugas akhir serta memberikan semangat dan hiburan di masa-masa sulit penyelesaian skripsi.
6. Segenap teman seperjuangan tugas akhir, Umar, Zilfa, Robby, Dee, Fitrah Haikal, Maulana, Dandy, Andika, Fandy, Alfian, Arian, Rendy dan Nublan yang telah membantu dan memberikan semangat serta hiburan di masa-masa sulit penyelesaian tugas akhir.
7. Serta pihak-pihak lain yang tidak disebutkan dan tanpa sadar telah menjadi inspirasi dan membantu penulis dalam menyelesaikan tugas akhir.

Penulis berharap semoga Tuhan Yang Maha Esa berkenan membalas segala kebaikan serta jasa dari semua pihak yang telah banyak membantu penulis dalam menyelesaikan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna. Oleh karena itu penulis mengharapakan segala bentuk saran serta masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberi manfaat bagi para pembaca dan semua pihak. Aamiin.

Makassar, November 2022

Penulis

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Indonesia memiliki keunggulan di bidang maritim yang tidak dimiliki oleh negara lain. Indonesia memiliki sekitar 17.499 pulau dengan luas total wilayah Indonesia sekitar 7,81 juta km<sup>2</sup>. Dari total luas wilayah tersebut sekitar 3,25 juta km<sup>2</sup> merupakan wilayah lautan dan 2,55 juta km<sup>2</sup> merupakan Zona Ekonomi Eksklusif (Pratama, 2020). Letak Indonesia berada di persimpangan antara dua benua dan dua samudera yaitu benua Asia dan benua Australia serta samudera Pasifik dan samudera Hindia sehingga menjadi persimpangan lalu lintas di dunia dan jalur penting perdagangan dunia (Rustam, 2016).

Indonesia memiliki 3 alur pelayaran yang disebut Alur Laut Kepulauan Indonesia (ALKI). ALKI I melintasi Laut Cina Selatan - Selat Karimata - Selat Sunda, ALKI II melintasi Laut Sulawesi - Selat Makassar - Lautan Flores - Selat Lombok, ALKI III melintasi samudera Pasifik - Selat Maluku, laut Seram - Laut Banda. ALKI II dan III menjadi rute yang ramai dilalui kapal (Malisan, 2019)

Selat Makassar merupakan salah satu Alur Laut Kepulauan Indonesia (ALKI) yang memiliki tingkat lalu lintas laut yang tinggi. Namun, banyak insiden kecelakaan laut yang sering terjadi di Selat Makassar. Kasus kecelakaan laut di Selat Makassar selalu ada setiap tahunnya. Hal ini terbukti pada november 2017, kapal penangkap ikan mengalami kecelakaan akibat diterjang gelombang tinggi disertai badai yang menghantam kapal hingga kapal terbalik (Syaefudin, 2017). Pada 13 Juni 2018, kapal tenggelam di Selat Makassar dikarenakan ombak dan angin kencang (Rahma & Purnomo, 2018). Pada tahun 2020, 6 anak buah kapal atau ABK terombang ambing di Selat Makassar selama 26 jam setelah kapal yang mereka tumpangi tenggelam dikarenakan gelombang tinggi disertai hujan dan angin kencang (Sucipto, 2020). Pada maret 2021, kapal dihantam gelombang tinggi yang langsung menenggelamkan kapal di Selat Makassar (Hilmansyah, 2021). Pada Mei 2022, kapal nelayan tenggelam di Selat Makassar karena dihantam ombak besar yang menyebabkan mesin armada mati (Ismail & Rosana, 2022).

Berdasarkan dari paragraf diatas bahwa faktor alam atau cuaca buruk merupakan permasalahan yang dianggap sebagai penyebab utama terjadinya kecelakaan laut di Selat Makassar. Permasalahan yang dialami biasanya adalah badai, gelombang yang tinggi, arus yang besar juga kabut yang menghalangi pandangan (Ridwan & Desha Sanjaya, 2022). Berdasarkan beberapa insiden yang terjadi di Selat Makassar, gelombang laut yang tinggi mendominasi tingkat kecelakaan laut di Selat Makassar.

Penelitian sebelumnya yang dilakukan oleh Manurung dkk dalam memprediksi gelombang laut di teluk Jakarta pada 6 titik menggunakan informasi angin sebagai input untuk regresi model. Hasil penelitian mereka menunjukkan bahwa 6 titik data angin memperoleh nilai *Correlation Coefficient* (CC) sebesar 0.94 dan nilai *Root Mean Squared Error* (RMSE) sebesar 0.05 (Manurung et al., 2020).

Oleh karena itu, penelitian ini mengusulkan sistem prediksi gelombang laut di Selat Makassar pada 3 titik yaitu Selat Makassar bagian selatan, tengah dan utara. Penelitian ini menambah 2 variabel dari penelitian sebelumnya yaitu suhu permukaan laut dan curah hujan. Hal ini dilakukan karena faktor meteorologi dan oseanografi dianggap secara tidak langsung dapat mempengaruhi gelombang laut.

Penelitian ini mengusulkan metode asosiasi untuk menemukan hubungan atau pola item variabel kecepatan angin, arah angin, suhu permukaan laut, dan curah hujan terhadap tinggi gelombang laut. Algoritma asosiasi yang akan digunakan ialah algoritma FP-Growth dikarenakan algoritma ini hanya melakukan 2 kali *scanning database* sehingga dapat menghemat waktu dan memori. Berdasarkan pola antara item yang dihasilkan oleh algoritma FP-Growth, digunakan untuk menyaring fitur yang dianggap relevan dalam proses prediksi. Pembangunan model prediksi menggunakan algoritma *support vector regression*. Hal ini dikarenakan algoritma *support vector regression* memiliki performa yang baik dalam memprediksi data dengan *noise* atau *outlier*. Selain itu, algoritma *support vector regression* juga menggunakan kernel *trick* yang memungkinkan algoritma *support vector regression* untuk memperluas kemampuan linier model dengan memodelkan data dalam dimensi tinggi.



## **1.2 Rumusan Masalah**

Adapun rumusan masalah dari penelitian ini ialah sebagai berikut:

- a. Bagaimana menemukan parameter yang paling tepat untuk digunakan dalam memprediksi tinggi gelombang laut di Selat Makassar?
- b. Bagaimana membangun model sistem prediksi untuk memprediksi tinggi gelombang laut di Selat Makassar ?

## **1.3 Tujuan Penelitian/Perancangan**

Adapun tujuan dari penelitian ini ialah sebagai berikut:

- a. Menemukan parameter yang paling tepat terhadap tinggi gelombang laut di Selat Makassar.
- b. Membangun model sistem prediksi tinggi gelombang laut di Selat Makassar

## **1.4 Manfaat Penelitian/Perancangan**

Adapun manfaat dari penelitian ini ialah sebagai berikut:

- a. Memberikan pengetahuan mengenai parameter yang paling berpengaruh untuk memprediksi gelombang laut
- b. Memberikan prediksi tinggi gelombang laut di Selat Makassar sehingga menjadi informasi untuk membantu keputusan berlayar

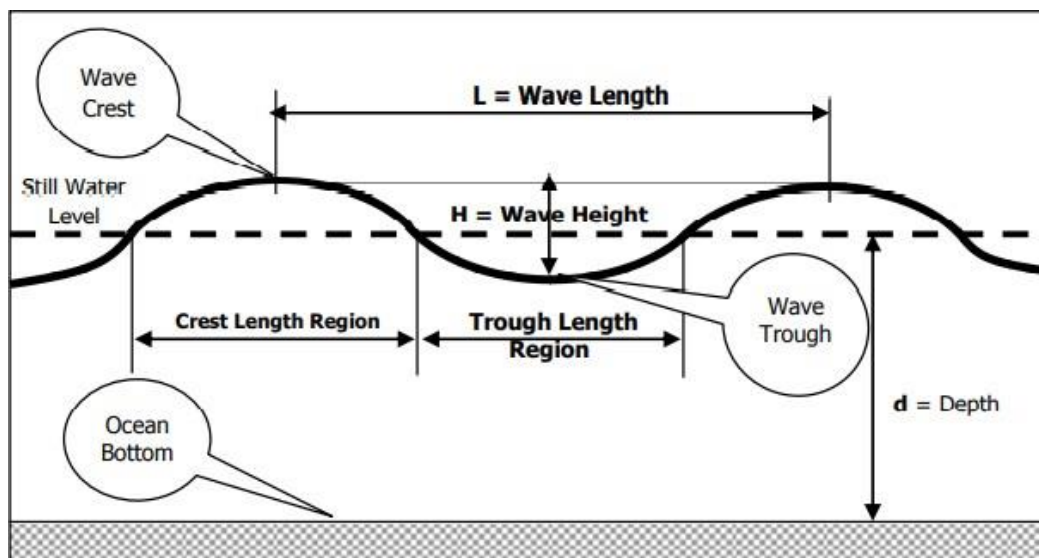
## **1.5 Ruang Lingkup/Asumsi perancangan**

- a. Data yang digunakan adalah data Selat Makassar
- b. Variabel-variabel penelitian terdiri dari
  - Kecepatan angin (m/s)
  - Arah angin (°)
  - Curah hujan (mm)
  - Tinggi Gelombang (m)
  - Suhu permukaan laut (°C)

## BAB II TINJAUAN PUSTAKA

### 2.1 Gelombang Laut

Gelombang laut merupakan pergerakan naik dan turun air laut yang disebabkan oleh gaya pembangkitnya. Berdasarkan gaya pembangkitnya, gelombang laut dapat dikategorikan menjadi beberapa jenis antara lain: gelombang laut yang dibangkitkan oleh tiupan angin dipermukaan laut disebut dengan gelombang angin, gelombang laut yang dibangkitkan oleh gaya tarik benda-benda langit terutama matahari dan bulan terhadap bumi disebut gelombang pasang surut, dan gelombang laut yang dibangkitkan oleh letusan gunung berapi atau gempa laut disebut gelombang tsunami, gelombang laut yang dibangkitkan oleh kapal bergerak disebut gelombang kapal dan sebagainya (Triatmodjo, 1999). Gambar 2.1 menunjukkan ilustrasi suatu gelombang.



Gambar 2. 1 Ilustrasi Gelombang

Keterangan :

$d$  = kedalaman laut (jarak antara muka air rerata dan dasar laut)

$H$  = tinggi gelombang =  $2a$  ( $a = \frac{1}{2} H$ )

$a$  = amplitudo gelombang

$L$  = panjang gelombang (jarak antara dua puncak gelombang yang berurutan)

$T$  = periode gelombang (interval waktu yang diperlukan oleh partikel air untuk kembali pada kedudukan yang sama dengan kedudukan sebelumnya)

$C$  = kecepatan rambat gelombang =  $L/T$

$\alpha$  = frekuensi gelombang =  $2\pi/T$

$k$  = angka gelombang =  $2\pi/L$

Gelombang angin merupakan gelombang yang paling dominan terjadi di permukaan laut dibanding jenis gelombang lainnya, baik ditinjau dari frekuensi kejadian maupun energinya (Kurniawan et al., 2012). Angin yang berhembus di atas permukaan laut memindahkan energinya ke air. Kecepatan angin menimbulkan tegangan pada permukaan laut, sehingga permukaan air yang semula tenang akan terganggu dan timbul riak gelombang kecil di atas permukaan laut. Semakin cepat angin, riak tersebut semakin besar dan apabila angin terus berhembus akan terbentuk gelombang. Semakin lama dan semakin kuat angin, semakin besar gelombang laut yang akan terbentuk (Hidayat, 2005).

Menurut Badan Meteorologi dan Klimatologi, klasifikasi tinggi gelombang laut berdasarkan skalanya ditunjukkan pada tabel 2.1 yaitu :

Tabel 2. 1 Kategori tinggi gelombang

Kategori	Interval (m)
Gelombang Tenang	0.1-0.5 m
Gelombang Rendah	0.5-1.25 m
Gelombang Sedang	1.25-2.50 m
Gelombang Tinggi	2.50-4 m
Gelombang Sangat Tinggi	4-6 m
Gelombang Ekstrem	6-9 m
Gelombang Sangat Ekstrem	9-14 m

Sumber : Badan Meteorologi dan Klimatologi (BMKG)

## 2.2 Angin

Angin merupakan udara bergerak yang disebabkan oleh tekanan udara di sekitarnya dan rotasi bumi. Angin bergerak dari tempat bertekanan udara tinggi menuju ke tempat bertekanan udara rendah. Angin juga terjadi karena ada perbedaan suhu. Angin yang mengenai energi panas, akan memuai sehingga menjadi lebih ringan dan tekanan udara turun karena udaranya berkurang. Oleh karena itu, angin berhubungan dengan sinar matahari. Panas yang datang dari matahari akan menimbulkan angin karena tekanan udaranya lebih rendah. Embusan angin akan mempengaruhi tinggi gelombang, kecepatan gelombang, serta panjang gelombang. (Hidayat, 2005). Pengelompokan angin sesuai Skala Beaufort (Sugianto, 2017) dibagi sesuai musim angin bertiup seperti berikut ini:

1. Musim barat (Januari, Februari, Dan Desember) dengan arah angin  $247,5^{\circ}$ - $337,5^{\circ}$ .
2. Musim peralihan (Maret – Mei Dan September – Nopember) dengan arah angin  $338^{\circ}$ - $44,5^{\circ}$ .
3. Musim timur (Juni-Agustus) dengan arah  $45^{\circ}$ - $112,5^{\circ}$

Tinggi dan periode gelombang yang dibangkitkan dipengaruhi oleh angin yang meliputi kecepatan angin ( $U$ ), lama hembus angin ( $D$ ) dan arah angin dari *fecth* ( $F$ ). Namun data angin yang dapat terukur yaitu arah angin dan kecepatan angin. Berikut merupakan detail kecepatan angin, arah angin, lama angin bertiup dan *fetch* :

- a. Kecepatan angin. semakin kencang angin, semakin besar pula gelombang yang terbentuk dan mempunyai kecepatan yang tinggi dan panjang gelombang yang besar (Putu & Setyawan, 2020). Kecepatan angin dianggap konstan jika perubahannya tidak lebih dari 5 knot (2,5 m/d) terhadap kecepatan rerata. Adapun distribusi kecepatan angin dengan skala Beaufort ditunjukkan pada tabel 2.2.

Tabel 2. 2 Distribusi kecepatan angin

Skala Beauford	Deskripsi	Kecepatan Angin (m/s)
1	Tenang	0-0.3 m/s
2	Sedikit Tenang	0.3-1.5 m/s
3	Sedikit Hembusan Angin	1.5-3.3 m/s
4	Hembusan Angin Pelan	3.3-5.5 m/s
5	Hembusan Angin Sedang	5.5-8 m/s
6	Sejuk	8-10.8 m/s
7	Hembusan Angin Kuat	10.8-13.9 m/s
8	Mendekati Kencang	13.9-17.2 m/s
9	Kencang	17.2-20.7 m/s
10	Kencang Sekali	20.7-24.5 m/s
11	Badai	24.5-28.4 m/s
12	Badai Dahsyat	28.4-32.6 m/s
13	Badai Topan	>32.6 m/s

Sumber : Skala Beaufort

- b. Arah Angin adalah arah dari mana angin berhembus dan dinyatakan dalam derajat arah (*direction degree*) yang diukur searah dengan arah jarum jam mulai dari titik utara bumi atau secara sederhana sesuai dengan skala sudut pada kompas (Prasetiyo et al., 2017). Arah angin masih bisa dianggap konstan apabila perubahannya tidak lebih dari 15°. Tabel 2.3 menunjukkan kategori arah angin (Tabler R.D, 1994)

Tabel 2. 3 Kategori arah angin

Kategori	Singkatan	Derajat
Utara	U	349°-11°
Utara Timur Laut	UTL	12°-33°

Timur Laut	TL	34°-56°
Timur-Timur Laut	TTL	57°-78°
Timur	T	79°-101°
Timur-Tenggara	TTG	102°-123°
Tenggara	TG	124°-146°
Selatan Tenggara	STG	147°-168°
Selatan	S	169°-191°
Selatan Barat Daya	SBD	192°-213°
Barat Daya	BD	214°-236°
Barat-Barat Daya	BBD	237°-258°
Barat	B	259°-281°
Barat-Barat Laut	BBL	282°-303°
Barat Laut	BL	304°-326°
Utara Barat Laut	UBL	327°-348°

Sumber : Tabler R.D, 1994

- c. Lama angin bertiup. Semakin lama angin bertiup semakin meningkat tinggi, kecepatan, dan panjang gelombang (Putu & Setyawan, 2020).
- d. *Fetch* merupakan daerah pembangkitan gelombang yang dibatasi oleh bentuk daratan yang mengelilingi laut. *Fetch* berpengaruh dalam pembangkitan tinggi gelombang dan periode gelombang dimana semakin panjang jarak *fetch* semakin besar ketinggian gelombang yang terjadi.

### 2.3 Suhu Permukaan Laut

Suhu permukaan laut Indonesia secara umum berkisar antara 26°C – 29°C dan variasinya mengikuti perubahan musim. Suhu permukaan laut dipengaruhi oleh panas matahari, arus permukaan, keadaan awan, *upwelling*, divergensi dan konvergensi terutama pada daerah muara dan sepanjang garis pantai. Perbedaan penerimaan radiasi matahari setiap wilayah menyebabkan perbedaan suhu, terkait dengan perbedaan letak geografis lintang. Suhu permukaan laut juga dipengaruhi

oleh kondisi meteorologi seperti penguapan, curah hujan, suhu udara, kelembaban udara dan kecepatan angin oleh karenanya suhu permukaan biasanya mengikuti pola musiman.

Suhu permukaan laut dipengaruhi juga oleh adanya El Nino dan La Nino. El Nino merupakan gejala penyimpangan (anomali) pada suhu permukaan Samudera Pasifik di pantai Barat Ekuador dan Peru yang lebih tinggi daripada rata-rata normalnya. Fenomena yang teramati adalah meningkatnya suhu permukaan laut yang biasanya dingin. Akibat adanya perubahan suhu permukaan laut di Pasifik ekuator, maka terjadi pula perubahan arah angin dan pergeseran kolom penaikan dan penurunan udara dari sirkulasi Walker dan ada saat yang bersamaan terjadi perubahan pola tekanan udara yang mempunyai dampak sangat luas dengan gejala yang berbeda-beda, baik bentuk dan intensitasnya. Kejadian ini dapat mempengaruhi iklim dunia. La Nina merupakan kebalikan dari El Nino ditandai dengan anomali suhu muka laut di daerah tersebut negatif / lebih dingin dari rata-ratanya (gejala penyimpangan/ anomali pada suhu permukaan Samudra Pasifik di pantai Barat Ekuador dan Peru berupa penurunan suhu). La Nina secara umum akan menyebabkan curah hujan di Indonesia meningkat (Kurniawan et al., 2012). El Nino dan La Nina turut berpengaruh terhadap dinamika oseanografi perairan Selat Makassar (Labania et al., 2019). Menurut Pusriskel (Pusat Riset Kelautan), kategori suhu permukaan laut ditunjukkan pada tabel 2.5

Tabel 2. 4 Kategori Suhu Permukaan Laut

Interval	Deskripsi
23.0°C-24.0°C	Ungu
24.0°C-26.0°C	Biru
26.0 °C-29.0°C	Hijau
29.0°C-30.0°C	Kuning
30.0°C-31.0°C	Orange
>31.0	Merah

Sumber : Pusat Riset Kelautan (Pusriskel)

## 2.4 Curah Hujan

Curah hujan merupakan jumlah air yang jatuh pada permukaan tanah selama periode tertentu bila tidak terjadi penghilangan oleh proses evaporasi, pengaliran dan peresapan, yang diukur dalam satuan tinggi. Tinggi air hujan 1 mm berarti air hujan pada bidang seluas 1 m<sup>2</sup> berisi 1 L atau  $100 \times 100 \times 0.1L = 1L$  (Kusmardiyanti & Yusuf, 2022). Menurut Badan Meteorologi dan Klimatologi, curah hujan di kategorikan seperti tabel 2.5

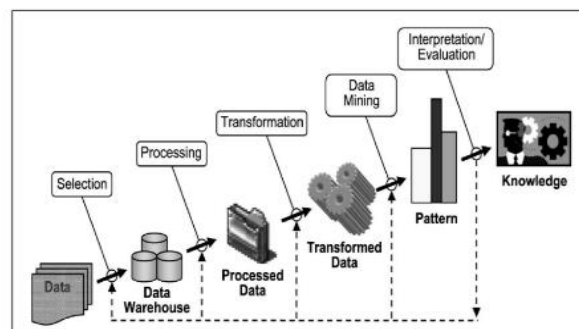
Tabel 2. 5 Kategori Curah Hujan

Interval	Kategori
5-20 mm	hujan ringan
20-50 mm	hujan sedang
50-100 mm	hujan lebat
>100 mm	hujan sangat lebat

Sumber : Badan Meteorologi dan Klimatologi (BMKG)

## 2.5 Knowledge Discovery in Database (KDD)

*Knowledge Discovery In Databases* (KDD) merupakan proses mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. Salah satu proses dalam KDD adalah *data mining*. Proses *Knowledge Discovery in Database* ditunjukkan pada Gambar 2.2 (Nofriansyah Dicky, 2015).



Gambar 2. 2 Proses Knowledge Discovery



Berdasarkan gambar di atas, proses Knowledge Discovery in Database (KDD) sebagai berikut :

### **1. Data Collection**

*Data collection* merupakan proses pengumpulan dan pengambilan data dari sumber-sumber data secara berkala (per hari, per jam, atau per sekian menit) atau secara real time (Kurniawan, 2021).

### **2. Data Selection**

*Data selection* merupakan proses seleksi data (pemilihan data) yang dianggap relevan/dibutuhkan kemudian selanjutnya data disimpan dalam berkas terpisah. Data hasil seleksi akan digunakan untuk proses data mining, memilih data apa saja yang dibutuhkan untuk diposes lebih lanjut (Nofriansyah, 2014).

### **3. Data Integration**

*Data integration* atau data integrasi merupakan proses menggabungkan data dari beberapa sumber data yang bertujuan untuk meminimalisir dan menghindari redundansi dan inkonsistensi untuk data yang baru. Dalam proses penggabungan data, hal yang perlu diperhatikan yaitu identifikasi entitas untuk memastikan atribut di kedua data yang akan digabungkan mengacu pada atribut yang sama (Han et al., 2012).

### **4. Data Cleaning**

Data yang diperoleh dari *data selection* memiliki isian yang tidak sempurna seperti *missing value*, data yang tidak valid atau kesalahan pengetikan yang dapat mengurangi nilai mutu dari data sehingga mengurangi akurasi dari hasil data mining sehingga diperlukan *data cleaning* (pembersihan data). Pembersihan data dilakukan untuk mengurangi jumlah data dan kompleksitasnya yang tidak memiliki mutu yang bagus. (Han et al., 2012).

### **3. Data Transformation**

*Data transformation* merupakan proses melakukan transformasi data atau konsolidasi data menjadi bentuk yang dapat diterima oleh algoritma data mining sehingga pola yang ditemukan dapat lebih mudah dipahami (Han et al., 2012). Adapun strategi untuk transformasi data sebagai berikut:

- *Smoothing*, memiliki fungsi menghilangkan noise dari data. Teknik

operasinya yaitu *binning*, regresi dan *clustering*

- *Attribute Construction*, membuat atribut baru dan ditambahkan dari kumpulan atribut yang diberikan untuk membantu proses mining data
- *Agregation*, melakukan operasi agregasi atau ringkasan
- *Normalization*, melakukan penskalaan data sehingga berada dalam rentang yang lebih kecil. Ada beberapa metode untuk melakukan normalisasi data yaitu *standard scaler* dan *minmaxscaler*. *StandardScaler* membuat mean = 0 dan menskalakan data ke varians unit. Sedangkan *MinMaxScaler* menskalakan dalam rentang [-1,1] atau [0,1]. Formulasi perhitungan normalisasi data menggunakan *MinMaxScaler* ditunjukkan pada persamaan (2.1)

$$x'_i = \left( \frac{x_i - x_{min}}{x_{max} - x_{min}} \right) \quad (2.1)$$

Keterangan :

$x'_i$  = hasil normalisasi data ke-i

$x_i$  = nilai data ke-i

$x_{min}$  = nilai data minimum dari keseluruhan

$x_{max}$  = nilai data maksimum dari keseluruhan

Sedangkan denormalisasi data adalah proses pengembalian data ke awal data sebelum dilakukannya normalisasi data untuk mendapatkan data asli. Proses denormalisasi dilakukan pada hasil akhir atau output peramalan. Perhitungan denormalisasi disajikan pada persamaan (2.2) sebagai berikut

$$x'_i = (y (\max - \min) + \min) \quad (2.2)$$

Keterangan :

y = Hasil output

min = Data minimum

max = Data maksimum

#### 4. *Data mining*

*Data mining* merupakan proses mengekstrak suatu data menjadi informasi (pengetahuan) dengan mengidentifikasi pola yang penting atau tersembunyi dari set data atau basis data besar. Data yang jumlahnya besar dan

menggunakan serta tidak berpola/tidak terstruktur diolah sehingga menemukan pola yang bermakna.

Secara umum, data mining dibagi menjadi 5 kelompok yaitu estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi. (Suntoro, 2018).

- Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan menggunakan record lengkap yang menyediakan nilai dari variabel prediksi (Nofriansyah, 2014).

- Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang (Nofriansyah, 2014)

Contoh prediksi dalam bisnis:

- i. Prediksi harga beras dalam tiga bulan yang akan datang.
- ii. Prediksi presentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dari teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi

- Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah (Nofriansyah, 2014).

- Pengklusteran

Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain (Nofriansyah, 2014).

- Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu (Nofriansyah, 2014).

## 5. Interpretation (Evaluation)

Tahap ini mencakup pada pemeriksaan pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 2.6 Association Rule

Dalam bidang keilmuan data mining, terdapat suatu metode yang dinamakan *association rule*. *Association rule mining* adalah suatu prosedur untuk mencari hubungan antar item dalam suatu data set yang ditentukan. Asosiasi dikenal sebagai salah satu teknik *data mining* yang menjadi dasar dari salah satu teknik *data mining* lainnya (Fauzy et al., 2016). Metode ini sering juga dinamakan dengan *market basket analysis*. Contoh aturan asosiasi dari analisis pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang konsumen membeli roti bersama dengan susu. Aturan asosiasi memberikan informasi dalam bentuk hubungan “*if-then*” atau “jika-maka” dan memeriksa semua kemungkinan hubungan *if-then* antar item dan memilih hanya yang paling mungkin (*most likely*) sebagai indikator dari hubungan ketergantungan antar item. Istilah *antecedent* untuk mewakili bagian “jika” dan *consequent* untuk mewakili bagian “maka” (Listriani et al., 2018). Metodologi dasar aturan asosiasi dijelaskan sebagai berikut (Listriani et al., 2016):

#### 1. Pembentukan Pola Frekuensi Tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam suatu *database*. *Support* yaitu seberapa banyak suatu item yang muncul dari keseluruhan transaksi. Nilai *support* adalah nilai penunjang atau persentase kombinasi sebuah item bersamaan dalam suatu *database*. Semakin besar nilai *support* menandakan semakin banyak data pendukung yang ditemukan dalam *database*. Nilai *support* sebuah item diperoleh dari persamaan 2.3 atau persamaan 2.4:

$$support(A) = \frac{\text{Jumlah transaksi yang mengandung A}}{\text{Total transaksi}} \quad (2.3)$$

Cara mencari nilai *support* dari 2 item :

$$support (A \cap B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Total transaksi}} \quad (2.4)$$

## 2. Pembentukan Aturan Asosiasi

Setelah seluruh pola frekuensi tinggi ditemukan , maka tahap selanjutnya adalah membentuk aturan asosiasi dengan melihat kombinasi item yang memenuhi syarat minimum dari nilai *confidence*. Nilai *confidence* adalah nilai keyakinan berupa kuatnya hubungan antar item yang didapatkan. Semakin besar nilai *confidence* menandakan semakin besar kemungkinan kombinasi item muncul secara bersamaan. Nilai *confidence* sebuah item diperoleh dari persamaan 2.5

$$Confidence P(B|A) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah transaksi mengandung A}} \quad (2.5)$$

Kedua parameter diatas digunakan untuk menentukan kekuatan suatu pola dan menemukan pola yang memenuhi syarat minimum untuk *support* (*minsup*) dan syarat minimum untuk *confidence* (*mincof*) (Priyana & Kardianawati, 2015).

## 3. Rasio Peningkatan (*Lift Ratio*)

*Lift Ratio* merupakan nilai yang menunjukkan keabsahan aturan yang terbentuk dalam proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. *Lift ratio* mengukur seberapa penting aturan yang telah terbentuk berdasarkan nilai *support* dan *confidence* yang telah didapatkan sebelumnya. Jika nilai *lift ratio* kurang dari atau sama dengan ( $\leq$ )1, maka hubungan sebab-akibat yang terjadi bersifat saling lepas satu sama lain. Sedangkan, jika nilai *lift ratio* lebih dari ( $>$ ) 1, maka hubungan sebab-akibat yang terjadi bersifat saling berhubungan satu sama lain dan dapat dikatakan kejadian tersebut bukan kebetulan dan akan berulang. Nilai *lift ratio* diperoleh dari persamaan 2.6:

$$Lift Ratio = \frac{\text{Confidence Antecedent}}{\text{Support Consequent}} \quad (2.6)$$

## 2.7 Algoritma *FP-Growth*

Algoritma *Frequent Pattern-Growth* atau yang biasa disebut dengan *FP-Growth* merupakan pengembangan dari algoritma apriori, sehingga pada algoritma *FP-Growth*, segala kekurangan dalam algoritma Apriori telah diperbaiki. *Frequent Pattern Growth (FP-Growth)* merupakan algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Algoritma *FP-Growth* hanya memerlukan 2 kali *scanning database* untuk menentukan *frequent itemset*. Struktur data yang digunakan untuk mencari *frequent itemset* dengan algoritma *FP-Growth* adalah perluasan dari penggunaan sebuah pohon *prefix*, yang biasa disebut adalah *FP-tree*. Dengan menggunakan *FP-tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent itemset* dari *FP tree* yang telah terbentuk dengan menggunakan prinsip *divide and conquer* (Samuel, 2008).

### 2.7.1 Pembangunan *FP-tree*

*FP-tree* merupakan struktur penyimpanan data yang dimampatkan. *FP-tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-tree*. Setiap transaksi yang dipetakan, mungkin ada transaksi yang memiliki item yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki item yang sama, maka proses pemampatan dengan struktur data *FP-tree* semakin efektif. Kelebihan dari *FP-tree* adalah hanya memerlukan dua kali pemindaian data transaksi yang terbukti sangat efisien (Samuel, 2008).

Misal  $I = \{a_1, a_2, \dots, a_n\}$  adalah kumpulan dari item dan basis data transaksi  $DB = \{T_1, T_2, \dots, T_n\}$ , di mana  $T_i$  ( $i \in [1..n]$ ) adalah sekumpulan transaksi yang mengandung item di  $I$ . Sedangkan *support* adalah penghitung (*counter*) frekuensi kemunculan transaksi yang mengandung suatu pola. Suatu pola dikatakan sering muncul (*frequent pattern*) apabila *support* dari pola tersebut tidak kurang dari suatu konstanta  $\xi$  (batas ambang *minimum support*) yang telah didefinisikan sebelumnya. Permasalahan mencari pola *frequent* dengan batas ambang *minimum support count* ( $\xi$ ) inilah yang dicoba untuk dipecahkan oleh *FP-Growth* dengan

bantuan struktur *FP-tree* (Samuel, 2008). Adapun *FP-tree* adalah sebuah pohon dengan definisi sebagai berikut:

- *FP-tree* dibentuk oleh sebuah akar yang diberi label *null*, sekumpulan upapohon yang beranggotakan item-item tertentu, dan sebuah tabel *frequent header*.
- Setiap simpul dalam *FP-tree* mengandung tiga informasi penting, yaitu label item, menginformasikan jenis item yang direpresentasikan simpul tersebut, *support count*, merepresentasikan jumlah lintasan transaksi yang melalui simpul tersebut, dan *pointer* penghubung yang menghubungkan simpul-simpul dengan label item sama antar lintasan, ditandai dengan garis panah putus-putus (Samuel, 2008). Sebagai contoh, tabel 2.6 merupakan data transaksi dengan *minimum support count*  $\xi=2$ .

Tabel 2. 6 Tabel data transaksi mentah

No	Transaksi
1	a,b
2	b,c,d,g,h
3	a,c,d,e,f
4	a,d,e
5	a,b,z,c
6	a,b,c,d
7	a,r
8	a,b,c
9	a,b,d
10	b,c,e

Sumber : Samuel, 2008

Kemudian dilakukan perhitungan frekuensi kemunculan tiap item dari semua item yang ada. Frekuensi kemunculan tiap item dapat dilihat pada tabel 2.7 berikut:

Tabel 2. 7 Frekuensi kemunculan tiap karakter

Item	Frekuensi
a	8
b	7
c	6
d	5
e	3
f	1
r	1
z	1
g	1
h	1

Sumber : Samuel, 2008

Setelah dilakukan pemindaian pertama di dapat item yang memiliki frekuensi di atas *support count*  $\xi=2$  adalah a,b,c,d, dan e. Kelima item inilah yang akan berpengaruh dan akan dimasukkan ke dalam *FP-tree*, selebihnya (r, z, g, dan h) dapat dibuang karena tidak berpengaruh signifikan. Tabel 2.8 menunjukkan kemunculan item yang *frequent* dalam setiap transaksi, diurut berdasarkan frekuensinya yang paling tinggi (Samuel, 2008).

Tabel 2. 8 Tabel data transaksi

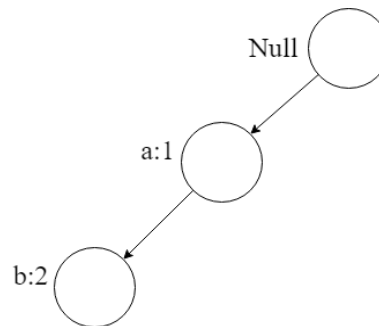
TID	Item
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	[a.b.c}
6	{a,b,c,d}



7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

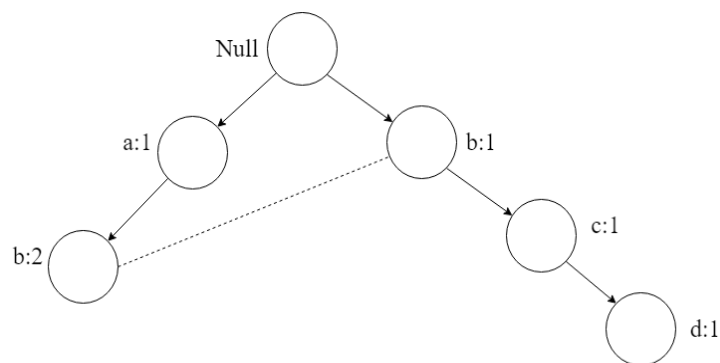
Sumber : Samuel, 2008

Gambar 2.3 memberikan ilustrasi mengenai pembentukan *FP-tree* setelah pembacaan TID 1.



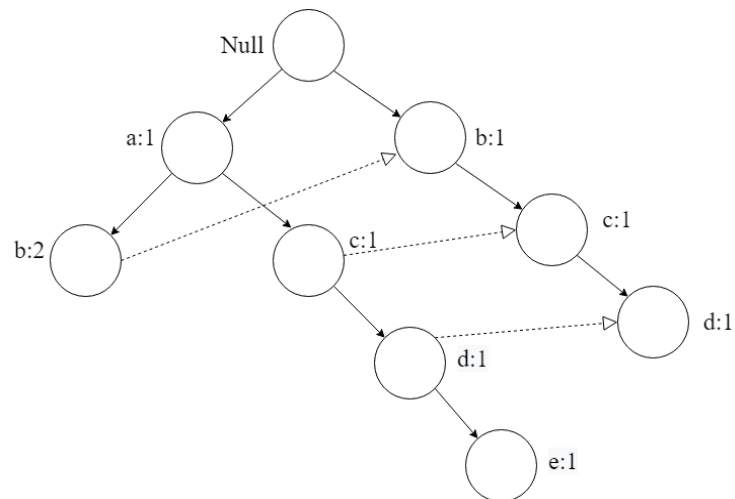
Gambar 2. 3 Hasil pembentukan FP-tree TID 1 (Samuel, 2008)

Gambar 2.4 memberikan ilustrasi mengenai pembentukan *FP-tree* setelah pembacaan TID 2.



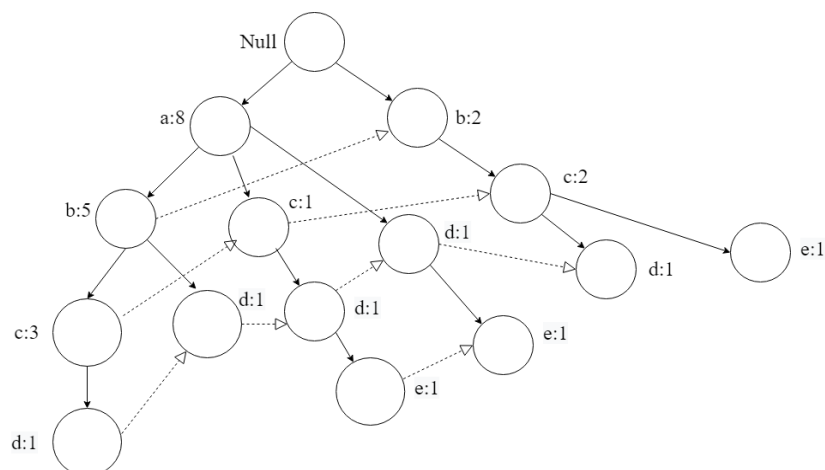
Gambar 2. 4 Hasil pembentukan FP-tree TID 2 (Samuel, 2008)

Gambar 2.5 memberikan ilustrasi mengenai pembentukan *FP-tree* setelah pembacaan TID 3.



Gambar 2. 5 Hasil pembentukan FP-tree TID 3 (Samuel, 2008)

Gambar 2.6 memberikan ilustrasi mengenai pembentukan *FP-tree* setelah pembacaan TID 10.



Gambar 2. 6 Hasil pembentukan FP-tree TID 10 (Samuel, 2008)

Diberikan 10 data transaksi dengan 5 jenis item seperti pada tabel 2.8. Gambar 2.3, Gambar 2.4, Gambar 2.5 dan Gambar 2.6 menunjukkan proses terbentuknya *FP-Tree* setiap TID dibaca. Setiap simpul pada *FP-tree* mengandung nama sebuah item dan *counter support* yang berfungsi untuk menghitung frekuensi kemunculan item tersebut dalam tiap lintasan transaksi (Samuel, 2008).

*FP-tree* yang merepresentasikan data transaksi pada tabel 2.6 dibentuk dengan cara sebagai berikut:

1. Kumpulan data dipindai pertama kali untuk menentukan *support count* dari setiap item. Item yang tidak *frequent* dibuang, sedangkan *frequent item* dimasukkan dan disusun dengan urutan menurun, seperti yang terlihat pada tabel 2.8
2. Pemindaian kedua, yaitu pembacaan TID pertama {a,b} akan membuat simpul a dan b, sehingga terbentuk lintasan transaksi  $\text{Null} \rightarrow a \rightarrow b$ . *Support count* dari setiap simpul bernilai awal 1.
3. Setelah pembacaan transaksi kedua {b,c,d}, terbentuk lintasan kedua yaitu  $\text{Null} \rightarrow b \rightarrow c \rightarrow d$ . *Support count* masing-masing *count* juga bernilai awal 1. Walaupun b ada pada transaksi pertama, namun karena *prefix* transaksinya tidak sama, maka transaksi kedua ini tidak bisa dimampatkan dalam satu lintasan.
4. Transaksi keempat memiliki *prefix* transaksi yang sama dengan transaksi pertama, yaitu a, maka lintasan transaksi ketiga dapat ditimpakan di a, sambil menambah *support count* dari a, dan selanjutnya membuat lintasan baru sesuai dengan transaksi ketiga (Gambar 2.4).
5. Proses ini dilanjutkan sampai *FP-tree* berhasil dibangun berdasarkan tabel data transaksi yang diberikan.

### 2.7.2 Penerapan Algoritma *FP-Growth*

Setelah tahap pembangunan *FP-tree* dari sekumpulan data transaksi, akan diterapkan algoritma *FP-Growth* untuk mencari *frequent itemset* yang signifikan (Samuel, 2008). Algoritma *FP-Growth* dibagi menjadi tiga langkah utama, yaitu :

#### 1. Tahap Pembangkitan *Conditional Pattern Base*

*Conditional Pattern Base* merupakan *subdatabase* yang berisi *prefix path* (lintasan awal) dan *suffix pattern* (pola akhiran). Pembangkitan *conditional pattern base* didapatkan melalui *FP-tree* yang telah dibangun sebelumnya.

#### 2. Tahap pembangkitan *conditional FP-tree*

Pada tahap ini, *support count* dari setiap item pada setiap *conditional pattern base* dijumlahkan, lalu setiap item yang memiliki jumlah

*support count* lebih besar sama dengan *minimum support count*  $\xi$  akan dibangkitkan dengan *conditional FP-tree*.

### 3. Tahap pencarian *frequent itemset*

Apabila *conditional FP-tree* merupakan lintasan tunggal (*single path*), maka didapatkan *frequent itemset* dengan melakukan kombinasi item untuk setiap *conditional FP-tree*. Jika bukan lintasan tunggal, maka dilakukan pembangkitan *FP-Growth* secara rekursif.

## 2.8 Machine Learning

*Machine Learning* (pembelajaran mesin) menyelidiki bagaimana komputer dapat belajar atau meningkatkan kinerjanya berdasarkan data. Program komputer secara otomatis belajar mengenali pola kompleks dan membuat keputusan cerdas berdasarkan data. Misalnya, masalah pembelajaran mesin memprogram komputer sehingga komputer dapat secara otomatis mengenali kode pos tulisan tangan di surat setelah mempelajari serangkaian contoh. Pembelajaran mesin adalah disiplin yang berkembang pesat (Han et al., 2012). Berikut merupakan beberapa pendekatannya :

### a. *Supervised Learning*

*Supervised Learning* adalah sebuah pendekatan untuk data yang memiliki informasi kelas atau label. Label merupakan variabel yang menjadi identitas dari tiap data dalam sekumpulan data, sehingga tujuan dari pendekatan ini adalah mengklasifikasi suatu data ke data yang sudah ada (Han et al., 2012).

### b. *Unsupervised Learning*

*Unsupervised Learning* adalah proses belajar yang tidak diawasi (*unsupervised*) sebab set data yang diberikan tidak memiliki label, sehingga keluaran dari *unsupervised learning* tidak dapat di tebak karena model tersebut belajar sendiri untuk menemukan informasi dari set data yang telah diberikan. Salah satu penerapan dari *unsupervised learning* adalah *association rules* (Han et al., 2012)

### c. *Semi-supervised Learning*

*Semi-supervised learning* adalah kombinasi antara *supervised* dan *unsupervised learning*. Data yang diberikan pada pendekatan ini memiliki

kombinasi antara data yang berlabel dan tidak berlabel.

#### d. *Active Learning*

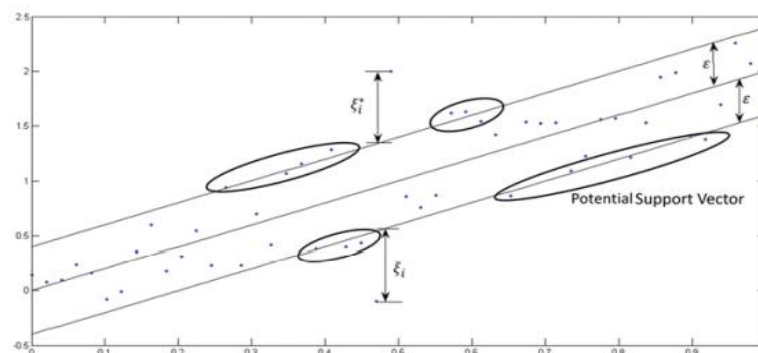
*Active learning* adalah kasus yang khusus atau spesial pada *machine learning* yang membuat sistem meminta peneliti atau pengguna secara interaktif memberi label data dengan keluaran yang diinginkan.

### 2.9 Peramalan (*Forecasting*)

Peramalan (*forecasting*) adalah sebuah seni dan ilmu untuk meramalkan kejadian di masa mendatang. Peramalan dapat dilakukan dengan melibatkan data masa lalu (historis) lalu memproyeksikannya ke masa depan dengan suatu bentuk model yang matematis (Apriani, 2016).

### 2.10 *Support Vector Regression (SVR)*

*Support vector regression (SVR)* merupakan pengembangan dari *Support Vector Machine (SVM)* untuk memecahkan permasalahan regresi. SVR bertujuan untuk menemukan sebuah fungsi  $f(x)$  sebagai suatu *hyperplane* (smola, 2002). *Hyperplane* terbaik dapat ditemukan dengan cara mengukur margin dengan *hyperplane* tersebut. Margin sendiri adalah jarak dari *hyperplane* dengan data yang terdekat. Data yang paling dekat dari margin disebut dengan *support vector* (Furi, Jordi, & Saepudin, 2015). Konsep dari SVR yaitu bagaimana caranya supaya semua data masuk ke dalam satu zona, dengan tetap meminimalisasi nilai *epsilon* ( $\epsilon$ ). *Epsilon* adalah jarak terjauh *support vector* dari *hyperplane*, kemudian disebut margin. Fungsi regresi yang mana sesuai dengan semua input data dengan sebuah *error* ( $\epsilon$ ) setipis mungkin.



Gambar 2. 7 SVR linier satu dimensi

Gambar 2.7 merupakan ilustrasi dari *support vector regression* dimana garis tengah atau *hyperplane* yang diapit oleh dua garis batas atau *epsilon*. Terlihat ada beberapa data poin yang dilingkari yang menjadi potensial *support vector* atau data poin yang dapat menjadi calon pembatas sehingga data poin masuk ke dalam satu zona dengan tetap meminimalisasi nilai epsilon. Kinerja pada SVR sendiri memiliki parameter yang mempengaruhi, yaitu:

- a. *Complexity* (C) yang menyajikan *trade-off* antara kompleksitas model dan jumlah sehingga penyimpangan yang lebih besar dapat ditoleransi. Untuk nilai C yang besar, pengoptimalan akan memilih *hyperplane* dengan margin lebih kecil jika *hyperplane* tersebut melakukan pekerjaan yang lebih baik untuk mengklasifikasikan semua poin pelatihan dengan benar. C adalah parameter untuk menentukan seberapa besar deviasi dari parameter *epsilon* yang masih dapat ditoleransi. Semakin besar nilai parameter ini, maka semakin besar nilai toleransi yang diperbolehkan (Aeni et al., 2020).
- b. *Epsilon* ( $\epsilon$ ) untuk mengontrol lebar zona *sensitive*, dimana  $\epsilon$  digunakan untuk menyesuaikan data latih dan digunakan sebagai pengatur batas kesalahan fungsi  $f(x)$ .
  - a. *Gamma* ( $\gamma$ ) menetapkan seberapa jauh *training* data yang berpengaruh dalam pembentukan *hyperplane*. Nilai *gamma* kecil berarti jauh dan sebaliknya.
  - b. *Degree* (d) berfungsi sebagai penentu dari dimensi yang akan dibuat oleh kernel polynomial, 1 berarti linear dan 2 berarti 2 dimensi array dan seterusnya. Semakin besar *degree*, semakin menyebabkan *overfitting*.
  - c. *Coef0* (c) atau *coefficient* mengatur seberapa besar model dipengaruhi oleh *degree*.

Bentuk umum dari *support vector regression* dapat dinyatakan pada persamaan (2.7)

$$f(x) = w^T \varphi(x) + b \quad (2.7)$$

Keterangan :

$\varphi(x)$  = titik di dalam feature space F, hasil dari pemetaan x di dalam input space

$f(x)$  = fungsi regresi

w = vektor bobot, yang mempunyai dimensi l

- b = bias  
 x = input atau vektor input

Sedangkan koefisien  $w$  dan  $b$  mempunyai fungsi untuk meminimalkan fungsi resiko, seperti pada persamaan (2.8) di bawah ini :

$$R = \min \frac{1}{2} \|w\|^2 + C \frac{1}{l} \left( \sum_1^l L_\varepsilon(y_i, f(x_i)) \right) \quad (2.8)$$

Keterangan :

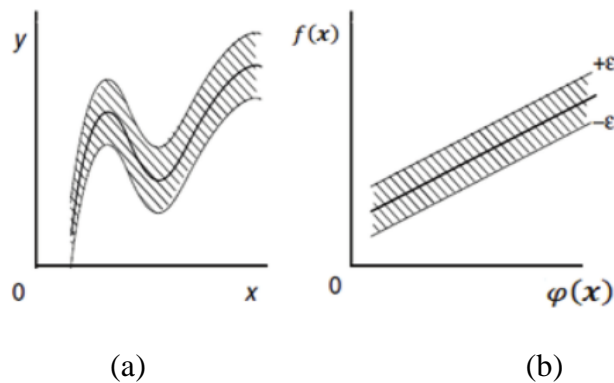
R = Risk Function (Fungsi Resiko)

$L_\varepsilon$  = loss function yang bertipe  $\varepsilon$ -insentive loss function

$\|w\|$  = normalisasi  $w$

$\varepsilon$  = epsilon yaitu deviasi atau derajat toleransi terhadap *error*

C = nilai penalti jika deviasi lebih besar dari batas *error*



Gambar 2. 8 Insensitive zone (a) original input space, dan (b) feature space

Pada Gambar 2.8 merupakan ilustrasi *insensitive zone* dan feature space pada *Support vector regression*.  $L_\varepsilon (y_i, f(x_i))$  menyatakan fungsi penalti untuk  $\varepsilon$ -insentive loss function, yang dikenai penalti jika *error*  $|y_i, f(x_i)|$  lebih dari atau sama dengan  $\varepsilon$ . Faktor  $\|w\|^2$  dinamakan regularisasi, meminimalkan  $\|w\|^2$  akan didapatkan fungsi setipis mungkin, sehingga bisa mengontrol *function capacity* (kapasitas fungsi). Faktor lain untuk menunjang ketelitian fungsi adalah memperhitungkan *empirical error* (kesalahan empirik) dengan menggunakan  $\varepsilon$ -insentive loss function. Koefisien  $\varepsilon$  ( $+\varepsilon$  dan  $-\varepsilon$ ) merupakan derajat toleransi terhadap *error*. Problem optimasi untuk kondisi *feasible* (pembatas yang layak) dapat diselesaikan menggunakan persamaan (2.9) sebagai berikut :

$$F = \min \frac{1}{2} \|w\|^2 \quad (2.9)$$

Dalam Persamaan (2.9) di atas, diasumsikan semua titik berada dalam *range* atau dalam margin dengan kondisi *Feasible* yaitu  $f(x) \pm \varepsilon$ . Pada kondisi *Infeasible* (ketidaklayakan pembatas), problem optimasi adalah ditunjukkan dengan adanya beberapa titik yang keluar dari *range* atau margin. Penyelesaian problem kondisi *Infeasible* ini adalah dengan menambahkan variabel *slack* ( $\xi$ ,  $\xi^*$ ), seperti Persamaan (2.10) sebagai berikut :

$$IF = \min \frac{1}{2} \|w\|^2 + C \sum_1^l (\xi_i + \xi_i^*) \quad (2.10)$$

Konstanta  $C > 0$ , berfungsi untuk menentukan *trade off* (tawar menawar) antara ketipisan fungsi  $f$  dan batas atas deviasi yang lebih kecil dari  $\varepsilon$ , yang masih bisa di toleransi.  $C$  adalah parameter untuk menentukan berapa besar deviasi dari parameter epsilon yang masih bisa ditoleransi, yaitu semakin besar nilai  $C$  maka semakin besar nilai toleransi yang masih diizinkan. Parameter *epsilon* ( $\varepsilon$ ) berfungsi untuk mengontrol lebar zona regresi dalam pemrosesan data. Semakin besar nilai epsilon maka estimasi regresi semakin datar mendekati regresi linier. Koefisien *epsilon* ( $\varepsilon$ ) dalam algoritma SVR adalah ekuivalen dengan tingkat akurasi *data training*. Nilai  $\varepsilon$  yang kecil berhubungan dengan nilai yang tinggi pada variabel *slack* dengan akurasi tinggi. Sebaliknya nilai  $\varepsilon$  yang besar atau tinggi berhubungan dengan nilai variabel *slack* yang kecil dan akurasi yang rendah. Penambahan variabel *slack* ini untuk mengatasi masalah pembatas margin yang tidak layak dalam problem optimasi.

Fungsi kernel untuk mentransformasikan data pada *input space* yang non-linear ke dalam ruang fitur yang dimensinya tinggi (*feature space*), untuk menemukan *hyperplane* yang optimal. Permasalahan pada prediksi IHK ini merupakan permasalahan yang memiliki pola nonlinier sehingga untuk mengatasi masalah ketidaklinieran tersebut digunakan Fungsi Kernel. Fungsi kernel yang digunakan pada metode *support vector regression* adalah Linear, Polynomial, *Radial Basis Function* (RBF)

## 2.11 Root Mean Square Error (RMSE)

*Root Mean Square Error* (RMSE) merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0), maka hasil prediksi akan



semakin akurat (Suprayogi dkk, 2014). Nilai RMSE dapat dihitung dengan persamaan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2} \quad (2.11)$$

Keterangan :

RMSE = nilai *root mean square error*

y = nilai aktual

$\hat{y}$  = nilai hasil prediksi

n = jumlah data