

**USULAN TUGAS AKHIR**

**1. IDENTITAS PENGUSUL**

Nama : Mohamat Ulin Nuha  
NRP : 5108100164  
Dosen Wali : Ir. F. X. Arunanto, M. Sc.

**2. JUDUL TUGAS AKHIR**

***“Implementasi Metode Bagging NN-SVM untuk Prediksi Kebangkrutan”***

**3. LATAR BELAKANG**

Dalam menjalankan bisnis, sebuah perusahaan bisa mendapat banyak keuntungan atau justru bisa mengalami kerugian. Di saat persaingan bisnis semakin ketat seperti yang terjadi sekarang ini, prediksi kebangkrutan adalah hal yang penting dilakukan agar nasib perusahaan dan karyawannya bisa ditentukan. Dengan kemampuan komputasi yang semakin memadai, prediksi kebangkrutan bisa dilakukan dengan metode *data mining* atau penggalian data berdasarkan variabel-variabel prediktor dari perusahaan tersebut.

Saat ini, sudah banyak metode yang ditawarkan untuk menyelesaikan permasalahan prediksi kebangkrutan. Salah satunya adalah metode yang berbasis *Nearest Neighbor Support Vector Machines* <sup>[1]</sup>. Metode *Nearest Neighbor Support Vector Machines* (NN-SVM) merupakan penggabungan dari metode *nearest neighbor* untuk mencari data latih (*training*) yang terdekat atau mirip dengan data uji (*testing*) dan *support vector machines* untuk menentukan klasifikasi atau prediksi. Pada implementasinya, algoritma *bagging* (*bootstrap aggregating*) akan diintegrasikan dengan metode NN-SVM untuk meningkatkan performa model.

Tugas Akhir ini bertujuan untuk mengimplementasikan metode *Bagging Nearest-Neighbor Support Vector Machines* untuk memprediksi kebangkrutan.

#### 4. TUJUAN TUGAS AKHIR

Tujuan dari Tugas Akhir ini adalah untuk mengimplementasikan metode *Bagging Nearest Neighbor Support Vector Machines* untuk memprediksi kebangkrutan.

#### 5. MANFAAT TUGAS AKHIR

Implementasi ini diharapkan dapat membantu memberikan prediksi kebangkrutan lebih awal bagi perusahaan, sehingga perusahaan bisa memilih keputusan yang tepat dalam menanggulangi resiko akan adanya kebangkrutan.

#### 6. RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam Tugas Akhir ini adalah:

1. Bagaimana membuat model *Bagging Nearest Neighbor Support Vector Machines (Bagging NN-SVM)*?
2. Bagaimana menguji model *Bagging NN-SVM* untuk memprediksi kebangkrutan perusahaan?

#### 7. BATASAN MASALAH

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki batasan, yaitu:

1. Dataset yang digunakan adalah *Wieslaw dataset*<sup>[2]</sup> dan *Australian credit dataset*.

#### 8. RINGKASAN TUGAS AKHIR

##### 8.1. Dataset

Pada implementasi Tugas Akhir ini digunakan *Wieslaw dataset* yang terdiri dari 30 *financial ratios* dan 240 kasus<sup>[2]</sup> dan *Australian credit dataset* yang diperoleh dari *UCI Repository of Machine Learning Databases*.

*Australian credit dataset* berisi 307 data pemohon kredit yang permohonannya layak disetujui dan 383 data pemohon kredit yang permohonannya tidak layak disetujui. Setiap data terdiri dari 6 atribut nominal, 8 atribut numerik, dan 1 atribut kelas (diterima atau ditolak).

Fitur pada dataset *Wieslaw* adalah sebagai berikut :

Tabel 1. Fitur Dataset Wieslaw

No.	Variabel prediktor	rasio finansial
X1	Cash/current liabilities	C/CL
X2	Cash/total assets	C/TA
X3	Current assets/current liabilities	CA/CL
X4	Current assets/total assets	CA/TA
X5	Working capital/total assets	WC/TA
X6	Working capital/sales	WC/S
X7	Sales/inventory	S/I
X8	Sales/receivables	S/R1
X9	Net profit/total assets	NP/TA
X10	Net profit/current assets	NP/CA
X11	Net profit/sales	NP/S1
X12	Gross profit/sales	GP/S
X13	Net profit/liabilities	NP/L
X14	Net profit/equity	NP/E
X15	Net profit/(equity + long term liabilities)	NP/EL
X16	Sales/receivables	S/R2
X17	Sales/total assets	S/TA1
X18	Sales/current assets	S/CA
X19	(365/receivables)/sales	R/S
X20	Sales/total assets	S/TA2
X21	Liabilities/total income	L/TI
X22	Current liabilities/total income	CL/TI
X23	Receivables/liabilities	R/L
X24	Net profit/sales	NP/S2
X25	Liabilities/total assets	L/TA
X26	Liabilities/equity	L/E
X27	Long term liabilities/equity	LTL/E
X28	Current liabilities/equity	CL/E
X29	EBIT (earnings before interests and taxes)/total assets	EBIT/TA
X30	Current assets/sales	CA/S

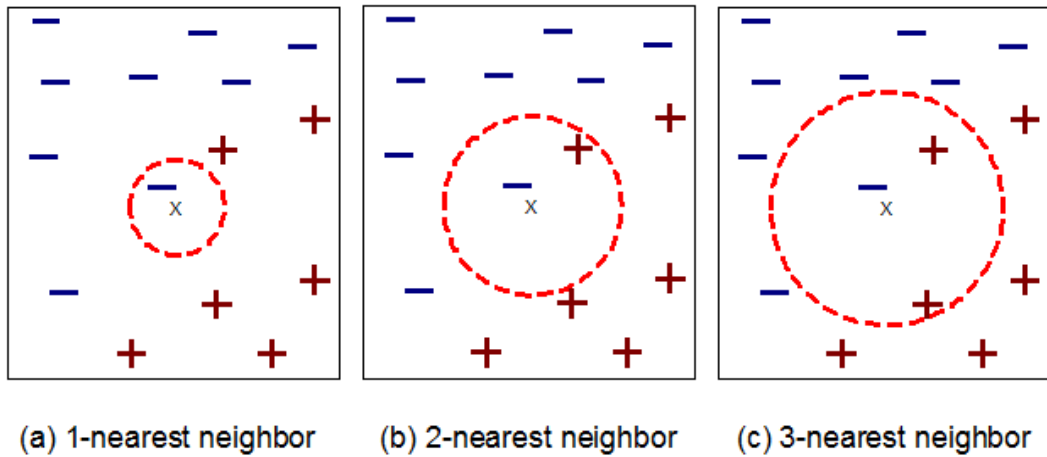
## 8. 2. K-Nearest Neighbor

*k-nearest neighbor* adalah metode klasifikasi objek berdasarkan data terdekat yang dimiliki oleh objek tersebut. Untuk melakukan klasifikasi dengan *nearest neighbor* dibutuhkan tiga parameter yaitu himpunan data, jarak antar data, dan nilai *k*, yakni jumlah data *nearest neighbor* yang akan dicari <sup>[3]</sup>.

Pada umumnya, jarak antar data dihitung dengan menggunakan *Euclidean distance*:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Untuk melakukan klasifikasi terhadap suatu data, langkah yang dilakukan adalah menghitung jarak antara data tersebut dengan data lain pada himpunan, memilih sejumlah  $k$  data terdekat (*k-nearest neighbor*) dari data tersebut, dan melakukan klasifikasi data berdasarkan *majority voting* pada label data *k-nearest neighbor* <sup>[3]</sup>.



Gambar 1. Ilustrasi klasifikasi dengan *k-nearest neighbor*

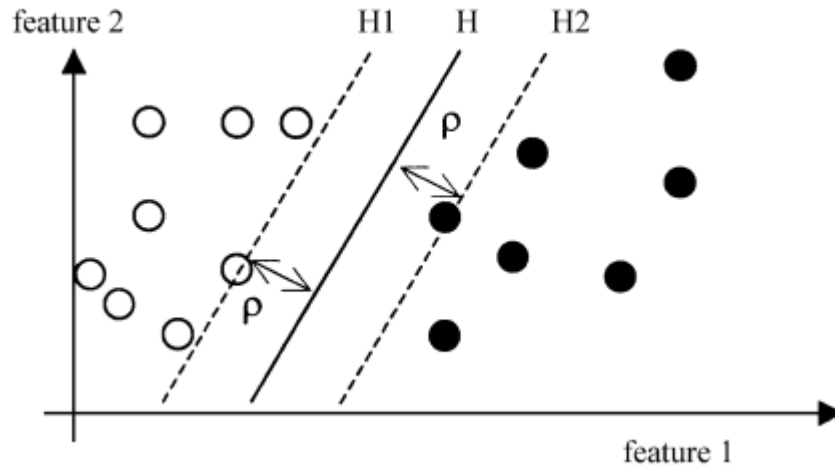
Pada Tugas Akhir ini, metode *nearest neighbor* digunakan untuk membentuk *base training set* dari dataset yang ada.

### 8. 3. Support Vector Machines

*Support vector machines (SVM)* adalah salah satu metode klasifikasi dan regresi. Ide utama dari SVM adalah menemukan *hyperplane* atau *decision boundary* yang bisa memisahkan dan mengklasifikasikan data menjadi dua kelas atau kategori <sup>[3-4]</sup>. *Hyperplane* yang dicari adalah *hyperplane* yang memiliki jarak maksimum terhadap titik-titik (*points*) terdekat pada *training set*. Titik-titik terdekat itu disebut dengan *support vectors* <sup>[4]</sup>.

SVM juga dapat diartikan sebagai algoritma linier pada *feature space* berdimensi tinggi yang berelasi secara nonlinier terhadap *input space*. Walaupun SVM merupakan algoritma linier pada *feature space* berdimensi tinggi, pada praktiknya, SVM tidak melibatkan komputasi pada *space* berdimensi tinggi

tersebut. Dengan penggunaan kernel, semua komputasi dilakukan pada *input space* [4].



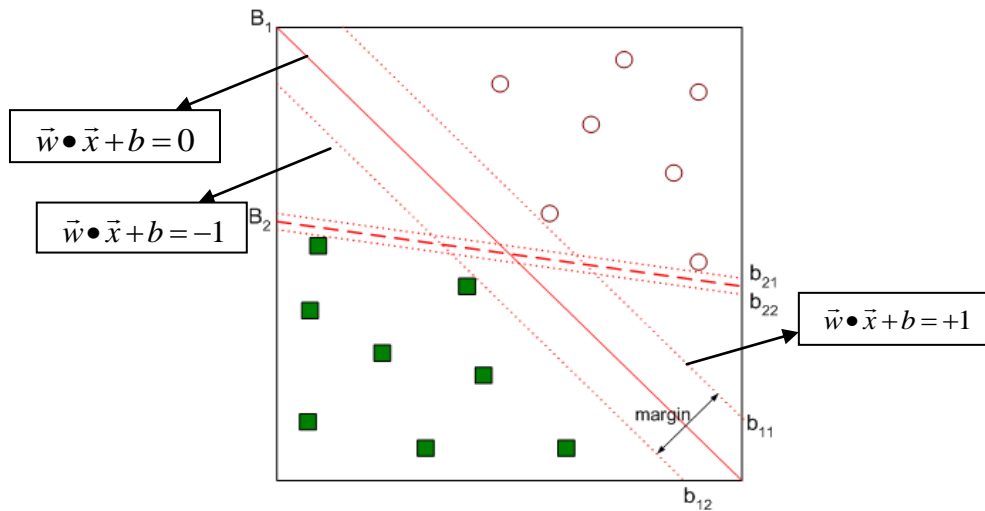
Gambar 2. *Hyperplane* H memisahkan *class point* terdekat dengan jarak yang sama.

Dalam membuat *hyperplane* pada *input space*, algoritma SVM terdiri dari dua tahapan. Tahapan pertama adalah memetakan *input space* ke dalam *feature space* dengan dimensi yang lebih tinggi dengan menggunakan *nonlinear mapping*. Selanjutnya SVM membuat *Maximal Margin Hyperplane (MMH)* pada *feature space* tersebut; MMH memaksimalkan margin atau jarak antara *hyperplane* dengan vektor terdekat yang menjadi anggota kelas yang berbeda [4].

Misalkan  $S$  adalah himpunan dari sejumlah  $l$  vector dengan  $x_i \in \mathbb{R}^n$ , ( $i = 1, 2, \dots, l$ ) pada *space*  $H$  berdimensi  $n$  (dimensi tinggi) yang disebut dengan *feature space*. Masing-masing vektor  $x_i$  menjadi anggota salah satu dari dua kelas yang diidentifikasi dengan label  $y \in \{-1, 1\}$ . Jika dua kelas tersebut dapat dipisahkan secara linier, maka ada sebuah *hyperplane*, yang didefinisikan dengan  $w \cdot x + b = 0$ , yang memisahkan  $S$  menjadi dua sisi dan menempatkan vektor-vektor yang memiliki kelas yang sama pada sisi yang sama [4].

Rumusan untuk *hyperplane* pada SVM adalah:

$$\begin{cases} \text{minimize } 1/2 \|w\|^2 \\ \text{with } y_i(w \cdot x_i + b) \geq 1 \quad (i = 1, 2, \dots, l) \end{cases} \quad (2)$$



Gambar 3. Ilustrasi klasifikasi dengan *support vector machines*

Pada kasus di mana himpunan  $S$  tidak bisa dipisahkan oleh *hyperplane* karena *partial overlapping* pada dua kelas yang berbeda, analisis atau rumusan sebelumnya bisa digeneralisasi dengan menambahkan variabel *slack* sejumlah  $l$  yang nilainya nonnegatif  $\xi = \xi_1, \xi_2, \dots, \xi_l$  sedemikian sehingga

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad (i=1, 2, \dots, l) \quad (3)$$

Solusi dari

$$\begin{cases} \text{minimize } 1/2 \|\mathbf{w}\|^2 + C \sum_i \xi_i & (i=1, 2, \dots, l) \\ \text{with } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i & (i=1, 2, \dots, l) \end{cases} \quad (4)$$

disebut dengan *Soft Margin Separating Hyperplane* (SMSH).

SVM pada Persamaan (4) memiliki batasan yakni ia tidak cocok untuk kasus distribusi data yang tidak seimbang (*imbalanced dataset*).

#### 8. 4. Bagging Nearest Neighbor Support Vector Machines

*Nearest Neighbor Support Vector Machines* (NN-SVM) adalah metode klasifikasi yang merupakan penggabungan metode *Nearest Neighbor* dan *Support Vector Machines*. Input dari sistem ini adalah *training set*, *testing set*, jumlah *base model* yang akan digunakan, dan jumlah *nearest neighbor* yang akan dihitung pada masing-masing data untuk membuat *base model*. Nilai *default* dari jumlah *base model* adalah 10 <sup>[1]</sup>.

Pengembangan model *NN-SVM* terdiri dari dua tahapan yaitu modul *training* dan modul *testing*. Modul *training* adalah modul untuk membangun model *NN-SVM* sedangkan modul *testing* adalah modul untuk klasifikasi data berdasarkan model *NN-SVM* yang telah dibuat.

Modul *training* terdiri dari tiga tahapan. Tahapan pertama adalah untuk setiap data pada *testing set* akan dicari *k-nearest neighbor*-nya pada *training set*. Pencarian *k-nearest neighbor* ini menggunakan algoritma *Euclidean distance*. Data *k-nearest neighbor* ini selanjutnya disebut dengan *base training set*. Jumlah *base training set* yang akan dibuat adalah 10. Selanjutnya, tahapan kedua, masing-masing *base training set* akan dibuat model SVM-nya. Model ini disebut *NN-SVM*. Tahapan terakhir adalah mengintegrasikan algoritma *bagging* dengan *NN-SVM* untuk menghasilkan *Bagging NN-SVM*.

Modul *testing* terdiri dari dua tahapan. Tahapan pertama adalah membuat *group decision* berdasarkan 10 model *Bagging NN-SVM*. *Group decision* ini dilakukan berdasarkan *majority voting* terhadap label dari *testing set*. Tahapan selanjutnya adalah melakukan prediksi atau klasifikasi dari masing-masing data *testing* berdasarkan hasil dari *group decision*.

## 9. METODOLOGI

Metodologi yang akan dilakukan dalam Tugas Akhir ini memiliki beberapa tahapan sebagai berikut:

### 1. Penyusunan Proposal Tugas Akhir

Tahap awal untuk memulai pengerjaan Tugas Akhir adalah penyusunan Proposal Tugas Akhir. Pada proposal ini, penulis mengajukan gagasan pembuatan sistem untuk memprediksi kebangkrutan perusahaan.

### 2. Studi Literatur

Tahapan Studi Literatur meliputi pencarian, pengumpulan, penyaringan, pembelajaran, dan pemahaman literatur yang berhubungan dengan algoritma metode *Bagging NN-SVM* serta pengumpulan *dataset*. Literatur yang digunakan dalam pengerjaan Tugas Akhir ini sebagian besar berasal dari internet berupa makalah ilmiah, tesis, artikel, materi kuliah, dan beberapa buku referensi.

### 3. Implementasi

Implementasi merupakan tahap untuk membangun sistem tersebut.

### 4. Pengujian dan Evaluasi

Pada tahap ini dilakukan uji coba terhadap sistem yang telah dibuat, mengamati kinerja sistem yang baru dibuat, serta mengidentifikasi kendala yang mungkin timbul.

### 5. Penyusunan Buku Tugas Akhir

Tahap terakhir dalam pengerjaan Tugas Akhir ini adalah penyusunan laporan yang memuat dokumentasi mengenai pembuatan serta hasil dari implementasi perangkat lunak yang telah dibuat.

## 10. JADWAL Pengerjaan Tugas Akhir

No	Tahapan	Bulan															
		Oktober				November				Desember				Januari			
1	Penyusunan Proposal																
2	Studi Literatur																
3	Implementasi																
4	Pengujian dan Evaluasi																
5	Penyusunan Buku Tugas Akhir																

## 11. DAFTAR PUSTAKA

- [1]. Li, Hui & Sun, Ji. 2011. Forecasting Business Failure: The Use of Nearest-Neighbour Support Vectors and Correcting Imbalanced Samples – Evidence from the Chinese Hotel Industry.
- [2]. P. Wieslaw. Application of Discrete Predicting Structures in An Early Warning Expert System for financial Distress. 2004, Ph.D. Thesis, Szczecin Technical University, Szczecin.
- [3]. Kumar, V. 2006. Introduction to Data Mining. New York: Pearson Addison-Wesley.
- [4]. Amendolia, S. R., Cossu, G., Ganadu M. L., Golosio, B., Masala, G. L., Mura, G. M. 2003. A Comparative Study of K-Nearest Neighbour, Support Vector Machine and Multi-Layer Perceptron for Thalassemia Screening.



## **LEMBAR PENGESAHAN**

Surabaya, 19 Oktober 2011

Menyetujui,

Dosen Pembimbing I,

Dosen Pembimbing II,

**(Isye Ariesianti, S.Kom, M.Phil)**  
**(NIP. 19780412 200604 2001)**

**(Yudhi Purwananto, S.Kom, M.Kom)**  
**(NIP. 19700714 199703 1002)**