ICTAS 2016
www.ictas.its.ac.id

The Analysis of Coronary Risk For Offshore Worker Using SVM Method

Farizi Rachman[1], Am Maisarah Disrinama[2], Kriselda Pranandaputri Istanto[3]
[1,2,3] Shipbuilding Institute of Polytechnic Surabaya, Indonesia

Corresponding author's e-mail: farizirachman@gmail.com

Abstract. Occupational Health and Safety is a Top Priority for offshore workers. As a result of occupational accidents Risks Should the disease is encountered by offshore workers hearts to do his job. Coronary Heart Disease (CHD) has become the first cause of death in Indonesia, especially workers. Many CHD patients obtained their risk factors. Major Risk Factors are lipid risk factors include the levels of cholesteroland trigliserida, because of the importance qualities substance encouraged the emergence of plaque in the arteries koroner. a study published outlining by journal The Lancet found the fact that the people who work more than 55 hours per week had a 13% risk for CHD than those who lack of it . On this research will perform analysis of CHD Risk Level Classification at offshore workers that already counted withFraminghram Method Score. The method used in thisresearchis Support Vector Machine (SVM). Based on the findings of the analysis, resulting the highest classification when using Value C = 100 is a function of the RBF kernel function ($\sigma = 3$) is equal to 67.85%. While currently using C = 10 highest classification using polynomial kernel functions with P = 1 and 71.71% classification accuracy.

Keywords: Coronary Heart Disease, OffshoreWorkers, Support Vector Machine, Classification Accuracy

1. Introduction

Physical health is a very important factor in supporting the productivity of work. The development of fast-growing industries require workers to work more competitively. Workers who suffered health problems will reduce physical abilities, thinking, and even cause accidents. Data from International Labour Organization (ILO) shows about 160 million workers become ill because of hazards in the workplace, and approximately 2.34 million died from the disease and work-related accidents [1]. Data from the National Center for Health Statistics, Centers for Disease Control and Prevention, and the United States Department of Health and Human Services [2] showed that in 2000 the highest death toll is based on the cause of the workers is coronary heart disease (710.760 cases), followed by cancer (553.091 cases) and stroke (167 661 cases).

Occupational Health and Safety is also a key priority for offshore workers. Workplace accidents due to illness is a risk that must be faced by offshore workers in doing their jobs. However, workers are important assets of the company, because it's physical and mental condition of employment must be observed. Periodic health checks need to be done to find out information regarding the health data of offshoreworkers. In accordance with the Government Policy Number 11th within year 1979 About Safety At Refining Oil And Gas stated that the head of the engineering required to pay attention to the

health of workers. It is important to do a health risk assessment of coronary heart disease in workers offshore in PT.X, as the embodiment of attention to the health of workers, especially the risk of coronary heart disease is very dangerous.

Coronary Heart Disease (CHD) has been the leading cause of death in Indonesia, especially workers. CHD patients mostly found a lot of their risk factors. The main risk factors or fundamental, namely lipids risk include cholesterol and trsigsilerida, because the importance of the properties of this substance in pushing the plaque in the coronary arteries. Various epidemiological studies have shown their circumstances and the nature of disorders that can accelerate the occurrence of coronary heart disease. Having more than one risk factor such as hypertension, diabetes mellitus, and obesity, it will have 2 times more risk of developing coronary heart disease [3, 4].

A study published in the journal The Lancet, found that those who worked more than 55 hours per week has CHD risk 13% higher than those who work less than that time. A system of work shift on offshore jobs in PT.X also a risk factor for CHD. It is related to the British Medical Journal study found a link between night shift workers with a 23% increased risk of heart attack, coronary heart disease by 24%, and 5% of strokes.

Based on the above presentation, it is important to conduct research on the analysis of the classification of the level of risk of CHD in offshore workers by using support vector machine (SVM).

SVM method is often used to classify health. Research using SVM ever done for the classification of breast cancer malignancy level by providing an accuracy up to 98% [6,7,8]. Application of SVM method for the prediction of coronary heart disease has also been performed with extremely high accuracy reached 90.27% [9].

This research will be analyze the classification level of CHD risk that has been calculated by the method Framinghram Score by using Support Vector Machine (SVM) on offshore workers in PT.X

2 Materials and Methods

2.1 Variable Data Source And Research

The research was conducted in the area of offshore petroleum companies PT.X with the object of research PT.X workers who work in the offshore area with the number of 93 people. The data used in this study are the variables included risk factors of Coronary Heart Disease (CHD) and a variable level of risk of coronary heart disease (CHD) are observed simultaneously. The level risk of coronary heart disease in previous offshore workers calculated by the method Framingharm Score. Research variables used in this study consisted of the response variable (y) and the predictor variable (x). The response variable in this study is the level of risk (CHD), which consists of three categories:

1. Low Risk (0)
2. Average (1)
3. High Risk (2)

Based on the literature formulated predictors variabelare summarized in Table 1. These variables consist of inspection data offshore workers amounted to 7 variables ((X1 - X8) as follows:

Table 1. Variable predictors research

No.	Variable	Category
1.	Age Number	1. < 34 yo 2. 35 – 44 yo 3. 45 – 54 yo 4. 55 – 64 yo 5. > 64 yo
2.	Total Cholesterol	1. <160 mg/dl 2. 160–190 mg/dl 3. 200-239 mg/dl 4. 240-279 mg/dl 5. ≥280 mg/dl
3.	HDL	1. ≥60 mg/dl 2. 50-59 mg/dl 3. 40-49mg/dl 4. <40 mg/dl
4.	Blood Pressure	1. <120 mmHg 2. 120-129 mmHg 3. 130–139 mmHg 4. 140-159mmHg 5. ≥160mmHg
5.	Diabetes Mellitus	1. Diabetes 2. No Diabetes
6.	Smoking	1. Smoking 2. No Smoking
7.	Work Period	1. < 5 years 2. ≥ 5 years
8.	Work Unit	1. PGF 2. SBU 3. LMO

2.2 Step Of Analysis

In conducting the study should be conducted proper analysis. Here's a step-by-step research:

1. To collect the primary data within offshore workers in accordance PT.X variable CHD risk factors (predictor variable) and the level of risk of CHD (the response variable).
   a. Translating medical language into a variable in Table 1
   b. Perform coding of data
2. Offshore workers descriptive analysis based CHD risk factors.
3. CHD risk level classification offshore worker with Support Vector Machine method. Here is an algorithm SVM method:
   a. Perform data transformation in accordance with the multi-class SVM method.
   b. Determining the function of separator method of multi-class one-on-many.
   c. Determine the values of the parameters C = 10 and C = 100, as well as determine the kernel function polynomial P = 1,2,3, and RBF with σ = 1, 2, 3
   d. Calculating the value of classification accuracy best
4. Calculation accuracy level of risk of CHD in offshore worker with SVM methods with a view confution matrix.

2.3 Support Vector Machine

Support Vector Machine (SVM) is a relatively new classification techniques. Scientists and practitioners have many apply these techniques in solving real problems in daily life. Both the problem gene expression, analysis, financial, weather, and others on medical field.

Support Vector Machine (SVM) perform a technique to find the function separators can separate the two sets of data from two different classes **.**This method is a machine learning method that works on the principle SructuralMinimazation Risk (SRM) with the goal of finding the best hyperplane that separates two classes in the input space**[10].**

Basically working with the principle of linear SVM clasifier, then developed to work on the case by using the concept of non-linear kernels on high dimension workspace[8]. In linear SVM classification is divided into 2 types, namely separable and non separable. Suppose X has a certain pattern, ie when x_i belongs to the class of the labeed x_i (target) =y_i+1 and y_i=-1.To that end, each label is denoted $y_i \in$ {-1, +1}, i = 1,2, ..l. So the data is a pair(x_1, y_1), (x_2, y_2), ..,(x_l, y_l). The data set is a pair of data for SVM. Support Vector Machine (SVM) can determine the pattern of generalization of x∈ X.

Basically, SVM is a method to classify the set of training vectors of the two classes (x_1, y_1), (x_2, y_2), ..,(x_l, y_l), with x∈ R^n, y{-1.1}. In the separation hyperplane with canonical form must satisfy constraint or constraint functions can be called (1)

$$y_i\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]\geq 1, i=1,...,l$$

Hyperplane that separates the data should minimize

$$\boldsymbol{\varphi}(w)=\frac{1}{2}\|\mathbf{w}\|^2 \qquad (2)$$

For the optimization of the equation 5,lagrange function used is

$$L(\mathbf{w},b,\alpha)=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{l}\alpha_i\{[(\mathbf{x}_i\cdot\mathbf{w})+b]y_i-1\} \qquad (3)$$

In this equation, the value is the lagrange multiplier function. The solution of the Lagrange function can be obtained by minimizing the primal variables L and L to maximize the (dual variables) and the completion of the following equation:

$$\hat{\alpha}=\arg\min_{\alpha}\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j(\mathbf{x}_i.\mathbf{x}_j)-\sum_{i=1}^{l}\alpha_i \qquad (4)$$

So the classification equation using the equation:

$$f(x)=sign\left(\mathbf{w}^T\cdot\mathbf{x}+\hat{b}\right) \qquad (5)$$

To address misclassification, formulations that have been made previously, will be expanded so that the data can be used non-separable. Previously optimization problem in both the objective function and constraints modified by following the slack variables ξ> 0, which owns a misclassification measure.Here is a constraint that has been modified for the case of non-separable:

$$y_i\left[(\mathbf{w}^T\cdot\mathbf{x}_i)+b\right]\geq 1-\xi, i=1,2,...,l \qquad (6)$$

Hyperplane or theoptimal separation is determined by the vector w, ie by minimizing the function:

$$\Phi(\mathbf{w},\xi)=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{l}\xi_i \qquad (7)$$

Where C is a parameter used to control the regulation of the relationship between the slack variables $\|\mathbf{w}\|^2$. The form of dual lagrange problem becomes:

$$\max_{\alpha} \mathbf{w}(\alpha) = \max(-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j (\mathbf{x}_i.\mathbf{x}_j) + \sum_{i=1}^{l}\alpha_i) \quad (8)$$

With constraints

$0 \le \alpha_i \le C, \quad i = 1,2,..l$

$$\sum_{i=1}^{l}\alpha_i y_i = 0 \quad (9)$$

And the solution of this problem:

$$\mathbf{w} = \sum_{i=1}^{l}\hat{\alpha}_i y_i \mathbf{x}_i \quad \text{dan} \quad \hat{b} = -\frac{1}{2}\mathbf{w}(\mathbf{x}_r + \mathbf{x}_s) \quad (10)$$

SVM also works on non-linear the data. Basically, non-linear classification of data having optimization equation:

$$\overline{\alpha} = \arg\min \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{l}\alpha_i \quad (11)$$

Value of K(x,y) is a kernel function that shows a non linear mapping in feature space. This equation gives a hard classifier on separating hyperplane in the feature space, the equation:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{SVs}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + \hat{\mathbf{b}}\right) \quad (12)$$

In the case of multiclass SVM can use several methods: one against all (SLA), one on one (SLU) and one optimization problem. The method used in this study is one against all (SLA). In this method for class classification, finding separators function k, where k is the number of classes.

2.4 Coronary Heart Disease

Heart conditions can also be influenced by other organs, because the heart is the only supplier of blood throughout the body, related to the transport function of cardiovascular, for example, the work of the heart will be affected by lifestyle (patterns of activity and diet) grown today are heart disease coronary [11].

The risk factors in the perspective of Occupational Health and Safety (K3) is known as a hazard, which can be defined as anything that could potentially cause harm in the form of injury or illness to workers and damage to property and the environment as well as the disruption of the company's image. Personal risk is what are the chances of potential hazard becomes a reality [12]. In this case, coronary heart disease is a multifactorial nature of disease [13]. Each risk factor contributing and each other will strengthen the risk of developing this illness. The risk factors consist of the risk factors that can be prevented and can not be prevented as well as additional risk factors.

CHD risk factors that can be prevented is smoking, hypertension (high blood pressure), cholesterol, and diabetes [11, 14, 15]. The second factor for CHD risk can not be prevented or can not be repaired

are age and gender [11, 14]. Besides these two factors there is also the additional risk factors are stress of work and the working environment [16, 17].

3. Result and Discussion

Based on the results of the risk assessment of coronary heart disease using the Framingham score in 2015, workers with high risk amounts to at most risk with 43 workers. While workers with average risk is equal to 34, and workers with low risk of risk amounted to 16.

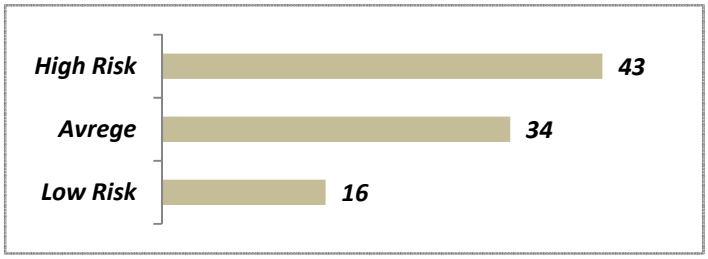

Figure 1. The level of risk of coronary heart disease in 2015 at PT.X

That condition is a condition that is very unsafe for the health of workers against coronary heart disease and should be done immediately control. Based on the data in Figure 1, Seen the number of workers to the risk of coronary heart disease High Risk still high. it shows that wellness programs work in the offshore area in PT.X that has not gone well.

Furthermore, SVM analysis on the level of risk of coronary heart disease in offshore work using two functions, namely polynomial kernel and Radial Basis Function. The level of risk of coronary heart disease in this study has been calculated using Framingharm Score.

On polynomial function using a parameter P as many as three types, namely P = 1, P = 2 and P = 3. While on the radial basis function $\sigma = 1$, $\sigma = 2$ and $\sigma = 3$. For the comparison value C is C = 10 and C = 100. By entering some values P in the kernel functions using polynomial and some value σ on the kernel function of radial basis function (RBF). So it can be compared to the best classification accuracy in the analysis of the level of risk of coronary heart disease in offshore work.

Table 2
Accuracy level classification SVM with C=10

Proportion	Polynomial			RBF		
	P=1	P=2	P=3	σ =1	σ =2	σ =3
50:50	65,21	67,39	65,21	47,8	60,8	52,7
70:30	50	53,57	60,71	53,75	60,71	**67,85**
80:20	42,8	57,14	28,57	57,14	42,85	42,85

Table 2 shows that the highest accuracy values contained in the data grouping using SVM Training and Testing 70:30, with radial basis function kernel function with value $\sigma = 3$. In this grouping an accuracy of 67.85%. While the smallest accuracy using polynomial kernel function using P = 3 with training and testing 50:50. In this grouping classification accuracy of only 28.75%.

Table 3
Tingkat Akurasi Klasifikasi SVM Dengan C=100

Proportion	Polynomial			RBF		
	P=1	P=2	P=3	σ =1	σ =2	σ =3
50:50	**71,71**	67,39	65,21	47,8	65,21	56,52
70:30	53,57	53,57	60,71	53,57	57,14	60,71
80:20	57,14	28,57	57,14	57,14	57,14	42,85

Table 3 shows that the highest accuracy for the classification level of risk of coronary heart disease in C = 100 contained in the data grouping using SVM Training and Testing 50:50, with a polynomial kernel function used with P = 1. In this grouping an accuracy of 71.71%. As for the smallest classification using polynomial kernel function with a value of P = 2.

4. Conclusion.

Based on the results of the classification level of risk of coronary heart disease in offshore work by using Support Vector Machine (SVM) concluded SVM classification using the highest value of C = 100 by using kernel functions radial basis function (RBF) with σ = 3 is equal to 67.85%. Meanwhile, when using the C = 10 highest classification using polynomial kernel function with P = 1 and the classification accuracy of 71.71%.

5. References

[1] International Labour Organizatation. The prevention occupational diseases-2 million workers killed every year [internet]. World Day for safety and health at work. 2013 [cited 2014 Jan 10]. Available from: http://www.ilo.org/wcmsp5/groups/public/edprotect/protrav/safework/documents/publication/wcms208226.pdf.

[2] National Alliance for Nutrition and Activity (NANA). Obesity and other diet-and inactivity-related diseases; national impact, costs, and solution. Nutrition policy [internet]. 2003 [cited 2011 May 11]. Available from:http://www.spencer.k12.ia.us/pages/hs_library/NANA advocatesnationalpolicies.pdf.

[3] Maulana, M. 2008. *Penyakit Jantung: Pengertian, Penanganan ,dan Pengobatan*. Penerbit Kata Hati, Yogyakarta.

[4] Soeharto, I. 2002. *Kolesterol dan Lemak Jahat Kolesterol, Lemak baik dan Proses Terjadinya Serangan Jantung dan Stroke* Cetakan kedua. PT Gramedia Pustaka Utama, Jakarta.

[5] Zahwardani, Diana. 2013. *Analisis factor risiko kejadian Penyakit Jantung Koroner di RSUP Dr Kariadi Semarang*. Jurnal Kedokteran Muhammadiyah Volume 1 Nomor 2 tahun 2013.

[6] Rachman, Farizi. 2012. *Klasifikasi Tingkat keganasan Breast Cancer menggunakan regresi logistic ordinal dan SVM*. Surabaya: ITS.

[7] Purnami, S.W. Abdullah Embong, Jasni M. Zain. 2009. *Application of Data Mining Technique using Best Polynomial Smooth Support Vector Machine in Breast Cancer Diagnosis*", International Conference on Robotics, Vision, Signal Processing and Power Applications(Rovisp 2009), Langkawi-Kedah, Malaysia.

[8] Purnami, S.W. and Abdullah Embong. 2008. *Smooth Support Vector Machine for Breast Cancer Classification,* The 4th IMT-GT 2008Conference on Mathematics, Statistics, and Their Applications (ICMSA08), Banda Aceh, Indonesia.

[9] Lusiyanti, Desy. 2014. *Perbandingan Metode Learning Vector Quantization (LVQ) dan Support vector machine*. Surabaya: ITS.

[10] Vapnik, V.N.(1999). *The Nature of Statistical Learning Theory*. New York: Springer

[11] Kurniadi, Helmanu. 2013. *Stop! Gejala Penyakit Jantung Koroner*. Yogyakarta : Familia
[12] Kurniawidjaja, Meily. 2011. *Teori dan Aplikasi Kesehatan Kerja* . Jakarta : UI- Press.
[13] Nadesul, Handrawan. 2008. *Sehat itu Murah*. Jakarta : Gramedia.
[14] Suryaningsih, Endang K. 2009. *Mengenal & Mencegah Penyakit Jantung, Kanker, Stroke*. Yogyakarta : Kirana Publisher.
[15] Patel, Chandra. 1987. *Fighting Heart Disease*. Great Britain : Dorling Kindersley Publisers Limited.
[16] Bronson, Ross. C et. Al. 1993. *Chronic Disease Epidemiology And Control. United State of America* : American Public Health Association
[17] Levy, Barry S. And David H. 1983. *Occupational Health Recognizing and Preventing Work Related Disease*. United State of Amerika : Little, Brown, and Company.

Acknowledgments
The author is very grateful to Surabaya Shipbuilding State Polytechnic that supports that research is going well. More specifically, the authors thank the PPNS P3K Laboratory, which already provides a place to carry out this research.