Risk Surplus Analysis in Credit Life Insurance Using Bayesian Method

Sukono, Khafsah Joebaedi, Nina Dora P., Endang Soeryana Hasbullah, Dwi Susanti
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
sukono@unpad.ac.id; khafsah.jbd@gmail.com; ninadora25@gmail.com; endang.soeryana@unpad.ac.id; dwi_susanti65@yahoo.com

Hasriati
Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Riau, Pekanbaru, Indonesia
hasriati.hasri@gmail.com

Abdul Talib Bon
Department of Production and Operations, University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

*T*his study discussed how to determine a surplus in credit life insurance. In this case, the risk model formed from the amount of claim and number of claims with the Bayesian method. Parameters of the number of claims and amount of claim are estimated with the Monte Carlo Markov Chain (MCMC) technique. The result of parameter estimation is applied to the collective risk model to determine premiums that need to pay by the insured person to insurer. In this study, the premium determined by the 4 calculation premium principle. The result shows the variance principle is the premium principle that given larger benefits for the insurance company.

Keywords
Credit life insurance, surplus process, Bayesian method, collective risk model, premium principle

1. Introduction
Credit Life Insurance (AJK) is an insurance program designed to protect the lives of creditors so that credit returns are on schedule. The benefit is to protect repayment/repayment of remaining credit (on schedule) if the debtor experiences an accident that causes death due to illness or accident in the period of insurance is still valid. The sum insured decreases with the decrease in the amount of the loan. Usually used for loans with fixed installments such as mortgage loans, cars, and the like. In credit life insurance, a surplus that grows every time is important. If the surplus-value is below or equal to zero, then the company has a chance of bankruptcy. The surplus is influenced by the number of premiums obtained and the number of claim payments out until a certain time. The important thing that must be done by insurance companies is to optimize the price of premiums so that the probability of bankruptcy for small companies (Dickson, 2005).

Some previous researches are related to insurance risk if the credit is as follows. Sukono et al. (2014), examines the probability of default on debt by a debtor in a financial services cooperative. The aim is to predict the risk of default (credit risk) prospective borrowers who apply for credit. An assessment is made of the loan application requirements document and field survey of prospective borrowers. In this credit valuation analysis is done using a logistic regression model, which parameter estimates are performed using genetic algorithms. Based on the results of an analysis of eight factors, it turns out that only six factors that have the potential to significantly incur the risk of default. These factors consist of: the number of dependents, the amount of savings, the value of collateral, monthly income, the realized credit limit, and the loan repayment period. Saputra et al. (2018), conducted a study on managing credit life insurance risk. Saputra said that for the management of credit life insurance risks insurance companies must be able to know their character to predict the likelihood of future losses. Credit life insurance risk characteristics can be studied through the claim distribution model. There are two ways to analyze the claim distribution model, namely the collective risk model and the individual risk model. Claims that arise when risks occur are individual claims, the accumulation of individual claims during the insurance period is an aggregate claim (collective). The aggregate claim model can be formed by a combination of the number and frequency of individual claims. Sukono et al. (2018.a), in their study, said that the risk model needs to be estimated by insurance companies, the purpose of which is to predict the magnitude of the risk of a claim and to determine the premium rates to the insured. This needs to be done to prevent future losses. In his study, the estimation of the risk model for motor vehicle insurance claims was conducted using the Bayesian method approach. In this research, the claim frequency is assumed to follow the Poisson distribution, and the magnitude of the claim is assumed to follow the Gamma distribution. Furthermore, the frequency distribution estimator and the amount of the claim are used to estimate the aggregate risk model, as well as to determine the average value and variance, which in turn is to determine the insurance premium rates to be paid by the insured.

Based on this description, this study aims to determine the credit life insurance surplus obtained by companies using the Bayesian method. The Bayesian method is used to estimate the model parameters of many claims and the size of the claims used for the calculation of the risk model and is useful for determining the price of premiums that must be paid by the insured to the insurance company.

2. Objects and Methods

The object of this research is the calculation of the surplus of credit life insurance risk. The research method used in this study is the Bayesian method.

2.1 Maximum Likelihood Method

The maximum likelihood method is a method that can be used in determining the estimated point of a parameter. For example, a random sample $X_1, X_2, ..., X_n$ of size *n*, derived from the opportunity density function $[f(x \mid \theta_1, ..., \theta_k)]$, the likelihood of the random sample is expressed by

$$L(\theta \mid x) = L(\theta_1, \theta_2, ..., \theta_k \mid x_1, ..., x_n) = \prod_{i=1}^{n} f(x \mid \theta_1, \theta_2, ... \theta_k) \tag{1}$$

2.2 Bayesian Method

Bayesian is one of the parameter estimation methods that use initial information about parameters or is called a prior distribution. According to Boldstad (2007) and Sukono et al. (2018.b), in Bayes estimation, after the sample information is taken and the prior has been determined, the posterior distribution is sought by multiplying the prior with the sample information obtained from the likelihood.

The posterior distribution is given by:

$$f(\theta \mid x) = \frac{f(\theta)L(\theta \mid x)}{\int_0^\infty f(\theta)L(\theta \mid x)d\theta}$$

According to Casadei et al. (2017), analytically, the function of the posterior density is obtained from the multiplication between prior and likelihood, which is stated as follows

$$f(\theta \mid x) \propto f(\theta)L(\theta \mid x) \tag{2}$$

2.3 Monte Carlo Markov Chain (MCMC)

The MCMC method is used to simulate samples from the posterior distribution. The most frequently used MCMC algorithm is the Gibbs Sampler. According to Gamerman, et al. (2006), the steps in the Gibbs Sampler algorithm are as follows:

a. Determine the initial value $\theta^{(0)} = \left\{\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_d^{(0)}\right\}$

b. If $X = (x_1, x_2, ..., x_n)$ is known, where X is the observation data, then for $j = 1, ..., J$, will be generated $\theta^{(j)}$ by repeating as follows:

$\theta_d^{(j)}$ from $f\left(\theta_d \mid \theta_1^{(j-1)}, \theta_2^{(j-1)}, ..., \theta_{d-1}^{(j-1)}, X\right)$

The Gibbs Sampler algorithm is a special case of Metropolis-Hastings, a pattern based on repetition or iteration of a conditional distribution, denoted as $f(\theta_i \mid \theta_{-i}, X)$, $i = 1, 2, ..., d$ which θ_{-i} shows all parameters θ except in θ_i.

2.4 Bayesian Inference Model Many Claims

Magnusson, and Hanna (2017), many claim N is a discrete random sample with the Poisson distribution expressed by $N \sim Poisson(\lambda)$. Based on equation (1) the likelihood function of many claims data, i.e.

$$L(n_i;\lambda) = \frac{1}{\prod_{i=1}^{t} n_i!} \lambda^{\sum_{i=1}^{t} n_i} e^{-t\lambda} \tag{3}$$

In Bayesian inference, many of the N claims with the Poisson distribution have the prior conjugate of Gamma distribution, with the FKP expressed as,

$$f(\lambda) = \frac{\beta_\lambda{}^{\alpha_\lambda} \lambda^{\alpha_\lambda - 1} e^{-\lambda \beta_\lambda}}{\Gamma(\alpha_\lambda)}, 0 \le \lambda < \infty \tag{4}$$

Based on equations (3) and (4), by ignoring equations that are not related to parameters a posterior equation is obtained that is proportional to,

$$f(\lambda \mid n_1, n_2, \ldots, n_t) \propto \lambda^{\alpha_\lambda + \sum_{i=1}^{t} n_i - 1} e^{-\lambda(\beta_\lambda + t)}$$

So it can be said that the posterior parameters of the gamma distribution, expressed as,

$$\lambda \mid n_i \sim Gamma(\alpha_\lambda + \sum_{i=1}^{t} n_i, \beta_\lambda + t).$$

2.5 Bayesian Inference Model Large Claim

Referring to Martín & Pérez (2009) and Lahcene (2020), the magnitude of claim X is assumed to be a continuous random sample that has a Lognormal distribution with parameters μ and σ^2 which are expressed with $X \sim LN(\mu, \sigma^2)$. Based on equation (1) the likelihood function of many claims data, i.e.

$$L(x_i;\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} \prod_{i=1}^{m} x_i^{-1} \exp\left[\sum_{i=1}^{m} \frac{-(\ln x_i - \mu)^2}{2\sigma^2}\right] \tag{5}$$

According to Sukono et al. (2017), the magnitude of claim X with Lognormal distribution has a prior gamma distribution conjugate, with FKP expressed as,

$$f(\sigma^{-2}) = \frac{\beta_\sigma{}^{\alpha_\sigma} (\sigma^{-2})^{\alpha_\sigma - 1} e^{-(\sigma^{-2})\beta_\sigma}}{\Gamma(\alpha_\sigma)}, 0 \le \sigma^{-2} < \infty \tag{6}$$

Based on equations (5) and (6), by ignoring equations that are not related to parameters σ^{-2}, a posterior equation is obtained that is proportional to,

$$f(\sigma^{-2} \mid x_i) \propto (\sigma^{-2})^{\alpha_\sigma + \frac{m}{2} - 1} e^{-(\sigma^{-2})\left(\beta_\sigma + \frac{\sum_{i=1}^{m} (\ln x_i - \mu)^2}{2}\right)}$$

So it can be said that the posterior parameters σ^{-2} of the gamma distribution, expressed as,

$$\sigma^{-2} \mid x_i \sim Gamma\left(\alpha_\sigma + \frac{m}{2}, \beta_\sigma + \frac{\sum_{i=1}^{m} (\ln x_i - \mu)^2}{2}\right) \tag{7}$$

Because the posterior parameters σ^{-2} of gamma are distributed, the parameters σ^2 of gamma inverse distribution are expressed

$$\sigma^2 \mid x_i \sim IG\left(\alpha_\sigma + \frac{m}{2}, \beta_\sigma + \frac{\sum_{i=1}^{m} (\ln x_i - \mu)^2}{2}\right) \tag{8}$$

2.6 Collective Risk Model

Referring to Bon et al. (2018), for example, the total number of claims for an insurance portfolio in the period *t* is notated (N_t), then the total number of claims (S_t), namely

$$S_t = \sum_{i=1}^{N_t} X_{t,i}$$

- Individual claims, $(X_{t,i})$, are non-negative random variables that are independent and of identical distribution.
- Many claims, (N_t), are random variables and are independent of claims $X_{t,i}$.

Expectations and variances from (S_t) expressed with,

$$E[S_t] = E(N_t)E(X_i) \tag{9}$$

$$Var(S_t) = E(N_t)Var(X_i) + [E(X_i)]^2 Var(N_t) \tag{10}$$

2.7 Calculation of Premiums

Refer to Magri et al. (2019) and Migon et al. (2006), Premiums are costs paid by the insured to the guarantor, the amount of which is determined by the guarantor. Premiums can be calculated based on the principle of premium calculation which consists of:

a. Pure Premium Principle

$$c = E[S]$$

b. The Principle of Expectation Value

$$c = (1+\theta)E[S]$$

which $\theta > 0$ shows the premium charge factor.

c. The Variance Principle

$$c = E[S] + \alpha Var[S]$$

d. Principle of Standard Deviation

$$c = E[S] + \alpha\sqrt{Var[S]}$$

which $\alpha > 0$ shows the premium charge factor.

2.8 Surplus Process

The surplus process at time t is defined as follows:

$$U_t = u + ct - \sum_{i=1}^{N_t} X_{t,i} \; ; \; ; t \geq 0$$

The surplus will be negative at any time. When this happens, it can be said that Ruin has occurred (Sidi et al, 2017).

3. Results and Discussion

3.1 Research Data

The study was conducted using secondary data, namely Bumiputera corporate credit life insurance claim data. In the calculation of risk surplus, the claim data obtained are grouped into two, namely the data of many claims and the number of claims. Many claims have discrete data, while claims have a continuous distribution.

3.2 Model Many Claims

3.2.1 Identification of Data Distribution of Many Claims

Using multiple claims data, determining the multiple claims distribution model is done by matching the distribution function graph using the help of EasyFit 5.6 software, the results are given in Figure 1.

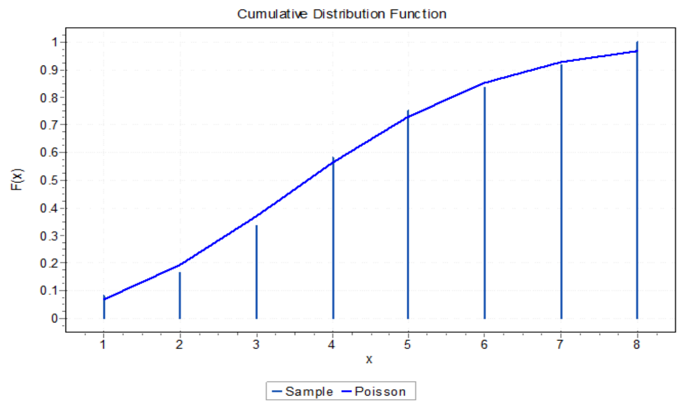

Figure 1. The distribution function curve of many claims data

Based on Figure 1, obtained a suitable distribution for many claims, namely Poisson distribution with parameter λ = 4.3333. To determine a more accurate fit distribution the Kolmogorov-Smirnov test was used. Based on the Kolmogorov-Smirnov distribution, the most suitable for many claims data is the Poisson distribution.

3.2.2 Bayesian Inference Model Many Claims

In this study low information prior was used, so the prior distribution for parameters λ was gamma with hyper-parameter $\alpha_\lambda = 0.001$ and $\beta_\lambda = 0.001$.

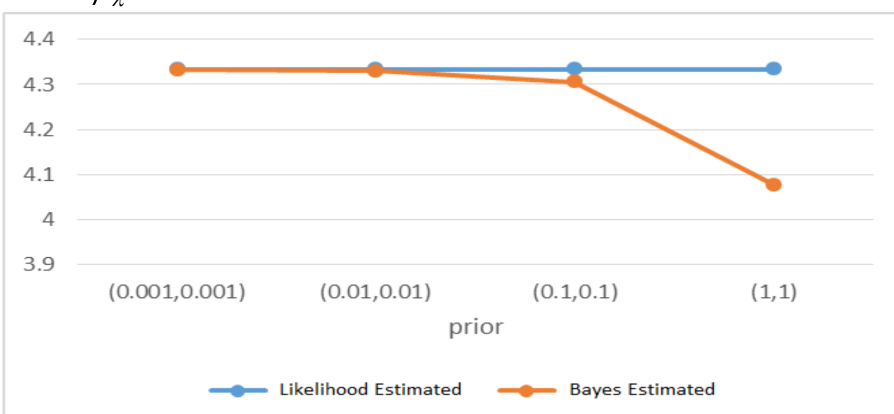

Figure 2. Comparison of estimated parameters with likelihood and Bayes

Because many claim data are known, as well as values of α_λ and β_λ have been determined, the posterior distribution of parameters λ can be expressed as,

$$\lambda \mid n_i \sim gamma(52.001; 12.001)$$

A summary of the statistics in the form of averages and variances of the parameters λ is stated in Table 1.

Table 1. Summary of statistical estimates of the Bayes method parameters λ manually

Parameter	Posterior α_λ	Posterior β_λ	Average	Variance	Standard Deviation
λ	52.001	12.001	4.333056	0.361058	0.600881

The sample simulation from the posterior distribution is carried out by the Monte Carlo Markov Chain (MCMC) method using the Gibbs Sampler algorithm. In this study, three chains are used to simulate the posterior distribution sample. The iteration for each chain is 10,000 iterations. Summary statistics with the openBUGS program yielding estimated Bayes parameters λ are shown in Table 2.

Table 2. Summary of statistical estimates of the Bayes parameter λ method with the OpenBUGS program

Parameter	Average	Standard Deviation	2.5% Percentile	97.5% Percentile
λ	4.331	0.6031	3.239	5.591

3.3 Large Claim Model

3.3.1 Identification of Data Distribution of Many Claims

The determination of the distribution model of many claims is done by matching the distribution function graph using the help of EasyFit 5.6 software, the results are given in Figure 3.

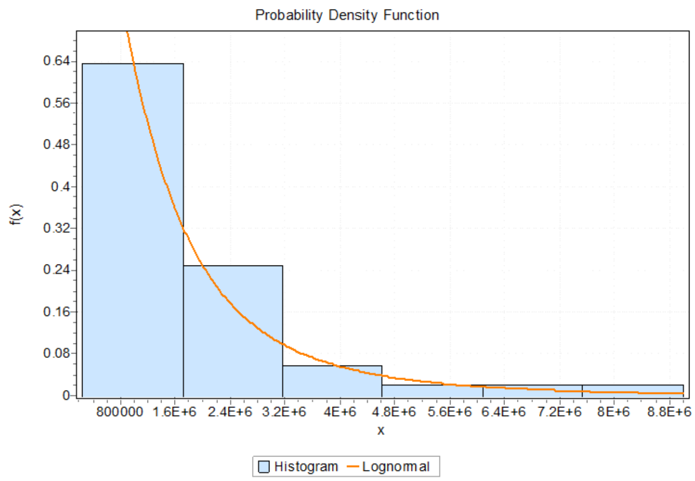

Figure 3. Histogram of the claim large data density function

Based on Figure 3, a suitable distribution is obtained for many claims, namely the Lognormal distribution with parameters σ = 0.9264 and μ = 13,872. To determine a more accurate fit distribution the Kolmogorov-Smirnov test was used. Based on the Kolmogorov-Smirnov distribution test the most suitable for large claims data is the lognormal distribution.

3.3.2 Bayesian Inference of Large Claim Model

In this study, priors of the gamma distribution parameters σ^{-2} with hyperparameter α_σ = 0.0001 and β_σ = 0.0001, because the estimated results using the Bayes method approach the estimated likelihood of the parameters σ^{-2}.

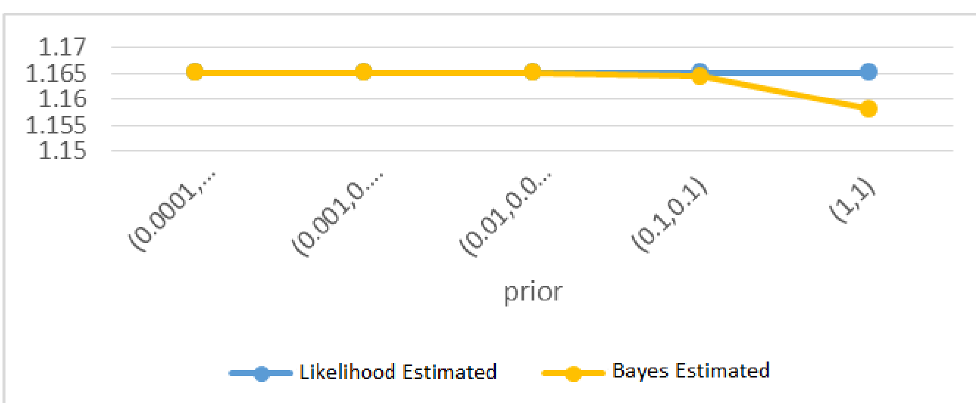

Figure 4 Comparison of estimated parameters σ^{-2} with likelihood and Bayes

A summary of statistics in the form of averages and variances of the parameters σ^2 is stated in Table 3.

Table 3. Summary of statistical estimates of Bayes parameters σ^2 obtained manually

Parameter	Posterior α_σ	Posterior β_σ	Average	Variance	Large Individual Claims
σ^2	26.0001	22.3139	0.8925541	0.0331937	1,575,883.1617

The sample simulation from the posterior distribution is carried out using the Monte Carlo Markov Chain (MCMC) method using the Gibbs Sampler algorithm. In this study, three chains were used to simulate posterior distribution samples. The iterations carried out on each chain were 10,000 iterations. The statistical summary with the openBUGS program yielded the estimated Bayes parameters σ^2 stated in Table 4.

Table 4. Summary of statistical estimates of Bayes parameters σ^2 with the OpenBUGS program

Parameter	$E(\sigma^2)$	Standard Deviation σ^2	2.5% Percentile	97.5% Percentile	*E(X)*
σ^2	0.8807	0.9338	0.7723	1.137	1,559,407.1030

3.4 Risk Collective Model

Based on equations (2.9) and (2.10), the expectations and variances obtained from the collective risk model are $E[S_t]$ = 6,753,792,164; and $Var[S_t]$ = 109174114439967x1011.

3.5 Calculation of Premium

The amount of premium to be paid by the insured to the insurance company is stated as given in Table 5.

Table 5. The amount of Credit Life Insurance Premiums

The Principle of Calculating Premiums	Premium Value
Pure Premium	562,816.014
Principle of Expectation Value	619,097.615
Principle of Variance	909.784,286,725x10^8
Principle of Standard Deviation	330,421,817,673.890

3.6 Calculation of Surplus

The surplus amount obtained by the insurance company is based on the premium calculation above, the results of the calculation of the surplus amount as given in Table 6

Table 6. Amount of Credit Life Insurance Surplus

The Principle of Calculating Surplus	Surplus
Pure Premium	-77,003,907.84
Principle of Expectation Value	5,392,356.565
Principle of Variance	136,467,643,103,372$\times 10^{11}$
Principle of Standard Deviation	3,964,978,054,386.710

4. Conclusion

Based on the analysis and discussion that has been done, it can be concluded that: based on the results of data processing, an exact distribution is obtained for many claims, namely Poisson distribution with parameters, and the amount of claims has a lognormal distribution with parameters μ and σ^2. Each of the estimated parameter values obtained using the Bayesian method is λ = 4,331 and σ^2 = 4,331. By using the Bayesian method the aggregate claims expectation is IDR 6,753,792.16 with a variance of IDR109,174,114,439,967. The average value and variance can be used by insurance companies as a reference in determining the value of premiums. In this study, obtained premium calculations that provide the largest surplus for insurance companies, namely the principle of variance.

Acknowledgments

Acknowledgments are conveyed to the Director of General of Higher Education of the Republic of Indonesia, and Chancellor, Director of the Directorate of Research, Community Engagement, and Innovation, and the Dean of the Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, who have provided the Basic Research Scheme Grant with Contract Number**: 1827/UN6.3.1/LT/2020.**

Reference

Boldstad, W.M. *Introduction to bayesian statistics,* (2th Edition). New Zealand: A John Wiley & Sons, Inc., Publication, 2007.

Bon, A.T., Ismail, M.I.A., Sukono, and Effendie, A.R. Collective Value-At-Risk (ColVaR) In Life Insurance Collection. *International Journal of Engineering & Technology,* 7 (3.7), 25-28, 2018.

Casadei, D., Grunwald, C., Kroninger, K., and Mentzel, F. Objective Bayesian analysis of counting experiments with correlated sources of background. *Journal of Applied Statistics*, January 20, 1-19, 2017.

Dickson, D.C.M. *Insurance risk and ruin.*Cambridge: Cambridge University Press, 2005.

Gamerman, et al., *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2ndEdition). London: Chapman & Hall/CRC. 2006.

Lahcene, B. Extended Lognormal Distribution: Properties and Applications. *World Scientific News,* 145, 16-30, 2020.

Magnusson, and Hanna. *A Risk Surplus Model using Hawkes Point Processes.* U.U.D.M. Project Report 2015:11. Uppsala University. 2015.

Magri, A., Farrugia, A., Valletta, F., and Grima, S. An analysis of the risk factors determining motor insurance premium in a small island state: the case of Malta. *International Journal of Finance, Insurance and Risk Management,* Volume IX, Issue 1-2, 63-85, 2019.

Migon, P. *Bayesian Analysis of A Health Insurance Model. Journal of Actuarial Practice.*13:62. 2006.

Martín, J. and Pérez, C.J. Application of a generalized lognormal distribution to engineering data fitting. *ResearchGate*. Departamento de Matemáticas, Universidad de Extremadura, Cáceres, Spain, 869-874, 2009.

Saputra, A., Sukono, and Rusyaman. Risk adjustment model of credit life insurance using a genetic algorithm. *IOP Conf. Series: Materials Science and Engineering,*332, 012-017, 2018.

Sidi, P., Mamat, M., Sukono, and Supian, S. Evaluation model *for* risk insurance premiums *of* building damage caused *by.* flood*:* case study *in* Citarum *watershed*, Southern Bandung, Indonesia. *Journal* of *Engineering* and *Applied Sciences 12* (17): 4420-4425, 2017.

Sukono, Sholahuddin, A., Mamat, M., and Prafidya, K. credit scoring for cooperative of financial services using logistic regression estimated by genetic algorithm. *Applied Mathematical Sciences*, Vol. 8, No. 1, 45 – 57, 2014.

Sukono, Suyudi, M., Islamiyati, F., and Supian, S. Estimation model of life insurance claims risk for cancer patients by using Bayesian method. *IOP Conf. Series: Materials Science and Engineering*, 166, 012-022, 2017.

Sukono, Riaman, Lesmana, E., Wulandari, R., Napitupulu, H., and Supian, S. Model estimation of claim risk and premium for motor vehicle insurance by using Bayesian method. *IOP Conf. Series: Materials Science and Engineering*,300, 012027, 2018.a.

Sukono, Aisah, I., Tampubolon, Y.R.H., Napitupulu, H., Supian, S., and Sidi, P. Estimated value of insurance premium due to Citarum River flood by using Bayesian method. IOP Conf. Series: Materials Science and Engineering 332, 012-047, 2018.b.

Biographies

Sukono is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung, Indonesia. Currently, as Chair of the Research Collaboration Community (RCC), the field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Khafsah Joebaedi is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung, Indonesia. The field of applied mathematics, with a field of concentration of applied statistics.

Nina Dora P.is Graduates of the Mathematics study program, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Bandung Indonesia. The field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Endang Soeryana Hasbullah a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung, Indonesia. The field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Dwi Susantiis a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung, Indonesia. The field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Hasriati is a lecturer in the Study Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Riau, Pekanbaru, Indonesia. The field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a Ph.D. in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He's bachelor degree and diploma in Mechanical Engineering which he obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM, and MIM.