International Journal of Education and Social Science Research

ISSN 2581-5148

Vol. 6, Issue.6, Nov-Dec 2023, page no. 107-116

To cite this article: Ismaini Zain, Iswari Hariastuti, Mario Ekoriano, Indra Murty Surbakti, Erma Oktania Permatasari, Adatul Mukarohmah and Nilam Novita Sari (2023). ANALYSIS OF UNINTENDED PREGNANCY FACTORS ON WOMEN OF REPRODUCTIVE AGE IN EAST JAVA USING LOGISTIC REGRESSION WITH SMOTE-N, International Journal of Education and Social Science Research (IJESSR) 6 (6): 107-116 Article No. 866, Sub Id 1354

ANALYSIS OF UNINTENDED PREGNANCY FACTORS ON WOMEN OF REPRODUCTIVE AGE IN EAST JAVA USING LOGISTIC REGRESSION WITH SMOTE-N

Ismaini Zain[1], Iswari Hariastuti[2], Mario Ekoriano[3], Indra Murty Surbakti[4], Erma Oktania Permatasari[5], Adatul Mukarohmah[6] and Nilam Novita Sari[7]

[1,5,6,7]Department of Statistics, Faculty of Science and Data Analytics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
[2,3,4]National Research and Innovation Agency, Jakarta, Indonesia

DOI: https://doi.org/10.37500/IJESSR.2023.6607

ABSTRACT

Unintended pregnancy is a condition where pregnancy occurs without the desire of want to have children. The high number of unintended pregnancies shows the need to research to find out the factors attributed to unintended pregnancy. Unintended pregnancy data used as a response variable includes imbalance binary data, which requires the use of logistics regression analysis. The imbalance of unintended pregnancy data causes a misclassification where a minority class sample can be classified as a majority class. One of the methods to overcome this imbalance is resampling. This research uses the Synthetic Minority Over-sampling Technique-Nominal (SMOTE-N) to overcome the imbalance. This technique synthesizes a new sample to balance the dataset by resampling the minority class sample. The data used in the research is the 2019 East Java data of the Accountability and Performance Survey. The sample is 8327 women of reproductive age. The variables which are expected to affect unintended pregnancy are age, education, occupation, residence, marital status, number of living children, and contraceptive knowledge. The best model obtained from the performance through accuracy, sensitivity, specificity, and G-mean. The results show the average accuracy between the model without imbalance treatment shows 89.7 % accuracy compared to only 65.3 % accuracy of the logistics regression model using SMOTE-N. However, the sensitivity of the model without imbalance treatment is lower than that using SMOTE-N. Moreover, the specificity and the G-mean show a not available value (N.A.), which indicates there is an imbalance that cannot classify data of the minority class sample. The results of the regression based on the Odds Ratio (OR) show that women aged 25-34, aged □ 35, have higher education, working, married, living in rural areas, have more than two children, and have good contraception knowledge are at high risk of having an unintended pregnancy.

KEYWORDS: Imbalanced Data, Logistics Regression, SMOTE, unintended pregnancy

International Journal of Education and Social Science Research

ISSN 2581-5148

Vol. 6, Issue.6, Nov-Dec 2023, p no. 107-116

INTRODUCTION

Unintended Pregnancy is a pregnancy that is not wanted by one or both parents. Unintended Pregnancy is a pregnancy condition without the desire to have children, a pregnancy that occurs not at the desired time, or an unplanned pregnancy at the time of conception [1]. Unintended Pregnancy is one of the most critical challenges facing the global public health system that harms women's personal lives, their families, and society [2]– [4].

According to research, among 208.2 million pregnancies that occur worldwide, as many as 38% (75 million) of these pregnancies are unintended pregnancies [1]. Most unintended pregnancies in developing countries are caused by not using any modern contraceptive [1]. In Ethiopia, 24.7% of unintended pregnancies were due to contraceptive failure, and 42% were due to not using contraception [1]. The lack of knowledge about contraceptives is one of the reasons why unintended pregnancies cannot be prevented [5].

Unintended pregnancy is a health problem and has an impact not only on women but also on women's families and society in general [2], [3]. This unintended pregnancy is also one of the main reasons for the number of abortions [1]. For women aged 15-24 years, As many as 50 million cases of unintended pregnancy out of 75 million cases caused abortion, of which 29 million cases of abortion occurred in unsafe conditions [6]. Unsafe abortion is also one of the leading causes of maternal death. Around 80,000 women die from unsafe abortions [7].

Unintended pregnancies have an impact on the parents, children, and society. One indication of the impact of Unintended Pregnancy is that it is one of the main reasons for the number of illegal and unsafe abortions which is one of the main causes of death and disability among mothers. Women aged 15-24 years have a higher risk of maternal death due to physical immaturity which is also exacerbated by these adolescents' lack of knowledge and experience [5]. In addition, due to the unpleasant feelings and emotions of the mother during pregnancy, the mother experiences more stress and nervous tension, and may even have effects such as depression, suicide, and a decrease in the mother's quality of life [6]. The impact of unintended pregnancy is also more felt by women aged 15-24 or teenagers, such as expulsion from school, rejection by family and environment, forced marriage, depression, and others [8]. In addition, women who give birth to children in their teens will have an impact on children such as developmental disabilities, behavior problems, and poor academic performance [1].

Many studies have been conducted on unintended pregnancies related to factors such as maternal age, father's age, education level of mother and father, economic status, employment status, pregnancy history, etc. Among them is a study conducted by [5] on women aged 15-24 years in Tanzania using regression. The results of this study indicate that age, education level, occupation, and marital status are factors that have a significant effect on unintended pregnancies in adolescents. In 2017, research on the risk factors for unintended pregnancy in Indonesia was conducted [9] using

health and demographic survey data. The results of this study indicate that maternal age, number of births, place of residence, and history of contraceptive use have a significant relationship with an unintended pregnancy. In contrast, employment status, education level, and family economic status significantly affect unintended pregnancy.

In this study, an analysis will also be carried out to determine the factors behind the occurrence of unintended pregnancies using several variables that are thought to have an effect, namely age, education, occupation, residence, marital status, number of living children, and contraceptive knowledge. One of the methods that can be used is logistic regression. Logistic regression is a statistical method used to predict an event where the response variable is binary (dichotomous) [10]. Logistic regression does not assume a linear relationship between the dependent and independent variables. The purpose of logistic regression is to predict the categorical response variables using the best model correctly. Logistic regression calculates the probability of success from the probability of failure, where the results of the analysis are transformed in the form of OR (Odd Ratio) [10].

This study uses secondary data from the Program Accountability Performance Survey (SKAP), which shows that the number of unintended pregnancies in East Java in 2019 was 855 out of 8327. Empirically, it shows that the data on the number of unintended pregnancies in East Java is not balanced. Analyzing unbalanced data becomes an interesting study because it needs to handle unintended pregnancy data so that the data becomes balanced. This study used the SMOTE-N method to overcome the imbalanced data. The SMOTE-N method is a development of the SMOTE method, which is a method that creates synthesis data in the minority class for data with nominal attributes [11], [12]. The difference between the SMOTE and SMOTE-N methods is the distance used [12]. Research on the classification of imbalanced data using SMOTE and logistic regression has been carried out by [13] using several datasets where the logistic regression estimates are compared using SMOTE and without SMOTE. The results showed that the accuracy produced without SMOTE was higher but produced biased results because minority data tended to be classified into the majority class. The logistic regression method with SMOTE produces higher AUC, sensitivity, and precision than without SMOTE. Another research by (Zain) compares the classification results without handling imbalance and by using SMOTE-N. The results of this study indicate that classification with imbalance handling using SMOTE-N produces better performance than without imbalance handling in terms of sensitivity, specificity, and G-Mean.

Pregnancy is a health problem that needs attention. So, determining the factors associated with unintended pregnancies is necessary to do. This study aimed to overcome the imbalanced unintended pregnancy data and classify unintended pregnancies by using the logistic regression method to obtain the factors that influence unintended pregnancies and the best model for the classification of unintended pregnancies.

International Journal of Education and Social Science Research

ISSN 2581-5148

Vol. 6, Issue.6, Nov-Dec 2023, p no. 107-116

METHODS

***Data Sources and Research Variables**:* The data used in this study is secondary data from the SKAP BKKBN module for Women of Reproductive Age (WUS) in East Java Province in 2019. The sample used is 8,327 women aged 15-49 years. The variables used in this study are as follows:

Y: Unintended Pregnancy (0: Yes, 1: Other)
X_1: Age ($0: < 25$ years, $1: 25 - 34$ years, $2: > 35$ years)
X_2: Education ($0: \leq$SLTP, $1: >$SLTP)
X_3: Occupation (0: Unemployee, 1:Employed)
X_4: Residence (0:Village, 1:City)
X_5: Marital Status (0:Married, 1:Other)
X_6: Number of Living Children ($0: \leq 2$, $1: > 2$)
X_7: Contraceptive Knowledge (0: No, 1: Yes)

***Research Design**:* In this study, the classification of unintended pregnancy will be carried out using the logistic regression method. Before classification, pre-processing of data is carried out to overcome missing values, after that, the data is divided into training and testing data using 5-fold cross-validation. The next step is to handle the imbalanced data by using SMOTE-N on the training data, after the data is balanced, the logistic regression method is applied to the data. The regression equation and OR will be obtained from the logistic regression model. After obtaining the model, the next step is to classify the testing data using the previously obtained model. Then the performance measurement on the results of the classification of the two methods is carried out using accuracy, sensitivity, specificity, and G-Mean. And the last is to compare the performance to get the best model for the unintended pregnancy classification.

Result: Table 1 shows the performance of logistic regression classification without imbalance handling and with imbalance handling using SMOTE-N. The average accuracy obtained by the model without handling imbalance is 89.70%. This is much higher than the average accuracy produced by the logistic regression model with SMOTE-N, which is only 65.30%. However, the sensitivity produced by the model without imbalance handling is lower than the model with SMOTE-N, and the specificity and G-Mean produced are Not Available, indicating an imbalance in the data because the model cannot classify the data in the minority class. The specificity and G-Mean produced by the logistic regression model with SMOTE-N increased, which means that the model with SMOTE-N overcame the imbalanced data and correctly classified the data in the minority class. Based on this classification performance, we can conclude that the model with SMOTE-N is better than the model without handling the imbalance. Therefore, we will use logistic regression with SMOTE-N to classify unintended pregnancy among women of childbearing age.

International Journal of Education and Social Science Research

ISSN 2581-5148

Vol. 6, Issue.6, Nov-Dec 2023, p no. 107-116

Table 1. Logistics Regression Classification Performance

	Accuracy	Sensitivity	Specificity	G-Mean
Logistic Regression	**0.897**	0.897	NA	NA
Logistic regression with SMOTE-N	0.653	**0.940**	**0.175**	**0.405**

Table 2 shows the performance of the logistic regression classification with SMOTE for each fold. Based on Table 2, it can be seen that the fold that produces the highest accuracy, sensitivity, specificity, and G-Mean values is fold 3, where the accuracy is 0.669, sensitivity is 0.949, specificity is 0.191, and the G-Mean is 0.425 so that the resulting model by fold 3 is the best.

Table 2. Logistics Regression Classification Performance With SMOTE-N

	Accuracy	Sensitivity	Specificity	G-Mean
Fold 1	0.658	0.933	0.167	0.395
Fold 2	0.639	0.942	0.171	0.401
Fold 3	**0.669**	**0.949**	**0.191**	**0.425**
Fold 4	0.667	0.940	0.179	0.410
Fold 5	0.633	0.938	0.166	0.394

From the logistic regression model generated by fold 3, the parameter estimates obtained can be seen in Table 3, where all variables show significant results, which means that the variables of age, education, job status, residence, marital status, number of children still alive, and contraceptive knowledge have a significant effect on the unintended pregnancy.

Table 3. Estimation of Logistics Regression Parameters

	B	Std. Error	P-value	Signifikansi
Constant	-2.412	0.123	0.000	*
Age ($\boldsymbol{X_1}$)(1)	1.812	0.105	0.000	*
Age ($\boldsymbol{X_1}$)(2)	1.652	0.103	0.000	*
Education ($\boldsymbol{X_2}$)(1)	0.291	0.043	0.000	*
Occupation ($\boldsymbol{X_3}$) (1)	0.752	0.043	0.000	*
Residence ($\boldsymbol{X_4}$)(1)	-0.476	0.044	0.000	*
Marital Status ($\boldsymbol{X_5}$)(1)	-0.958	0.083	0.000	*
Number of Living Children ($\boldsymbol{X_6}$)(1)	1.642	0.056	0.000	*
Contraceptive Knowledge ($\boldsymbol{X_7}$)(1)	0.280	0.076	0.000	*

*: significant at the level of 5%; -: not significant

International Journal of Education and Social Science Research

ISSN 2581-5148

Vol. 6, Issue.6, Nov-Dec 2023, p no. 107-116

Based on the parameter estimates obtained in Table 3. Then the equation can be written as follows.

$$Logit\left[g(x)\right] = -2.412 + 1.812\,X_1 + 1.652\,X_{1(1)} + 0.291\,X_2 + 0.752\,X_3 - 0.476\,X_4 - 0.958\,X_5 + 1.642\,X_6 + 0.280\,X_7$$

Table 4. OR and Confidence Interval OR of unintended pregnancy

	OR	CI low (2.5)	CI High (97.5)
Age ($\boldsymbol{X_1}$)(1)	6.125	5.005	7.544
Age ($\boldsymbol{X_1}$)(2)	5.216	4.277	6.404
Education ($\boldsymbol{X_2}$)(1)	1.338	1.229	1.457
Job Status ($\boldsymbol{X_3}$) (1)	2.121	1.949	2.309
Residence ($\boldsymbol{X_4}$)(1)	0.621	0.57	0.676
Marital Status ($\boldsymbol{X_5}$)(1)	0.384	0.325	0.451
Number of Children Still Alive ($\boldsymbol{X_6}$)(1)	5.166	4.632	5.771
Contraceptive Knowledge ($\boldsymbol{X_7}$)(1)	1.322	1.139	1.536

Discussion: Table 4 shows the OR and confidence interval for unintended pregnancy. This OR value can be used to interpret the model. The OR value can be used to express a relationship between categorical variables to determine the measure of risk between predictor variables and response variables. In Table 4. the OR value for the variable age 25-34 years is 6.125, which means that women with an age range of 25-34 years have a risk of having adverse events of 5,005 times compared to women with an age range of 25 years. Likewise, women in the 35-49 years age group with an OR value of 5,216 indicate that the risk of a 35-year-old woman having adverse events is 4,277 times compared to a 25-year-old woman. This shows a different result from the study conducted by [6], which showed that women in the age group of 20-34 years and 35 years had a lower risk of adverse events than women in the age group of 25.

Women with higher education have a higher risk of having an unintended pregnancy than women with low education. The results of this study are in line with the research by [14] which shows the same results, women who have a higher education have a higher risk of having unintended pregnancies compared to women who have low education. However, research from [15] shows the opposite results where the women's risk of having unintended pregnancy is higher when they have low education.

For Occupation variables, women who work have an OR of 2.121, which means that women who work have a 2,121 risk of having unintended pregnancy compared to women who do not work. Different results are shown by (5, 9, 16), which shows that women who do not work have a higher risk of having unintended pregnancies than women who work.

Women who live in urban areas have a higher risk of having unintended pregnancies than women who live in rural areas. The results of this study are in line with research from [17]–[19] which shows that women who live in urban areas have a higher chance of having unintended pregnancies than women who live in rural areas. But research from [9], [20] shows the opposite result where women who live in villages have a higher risk.

Single/divorced women have an OR of 0.384, which means that single/divorced women have a lower risk of having unintended pregnancy than married women. This study shows different results from the study [1], [5], [21] where single/divorced women at high risk have a higher risk of having unintended pregnancy than married women.

The variable number of living children shows the same results as research [14] which shows that women who have children who are still alive > 2 have a higher chance of having unintended pregnancy than women who have children living ≤ 2.

Furthermore, for the contraceptive knowledge variable, the results of this study indicate that women who have good knowledge of contraception are at risk of having unintended pregnancy by 1,322 times compared to women who have less knowledge of contraception, while [1] indicates that women who have good knowledge of contraception have an opportunity to having unintended pregnancy lower than those who have less knowledge of contraception.

The results of other studies in Pakistan explain that unwanted pregnancy generally occurs in women with low education who do not know how to have safe sexual intercourse, for example they do not know how to use contraceptives to prevent pregnancy.[22]. High prevalence of unwanted pregnancies and important factors especially those related to knowledge of family planning. An integrated national family program that provides contraceptive services, especially modern methods, to women during the pre-conception period, especially long-term contraceptive methods. Unfortunately, the results of the 2017 Indonesian Demographic and Health Survey (IDHS) show that 64% of Women of Reproductive Age (WUS) have used contraception. But unfortunately only 13% of women use MKJP such as IUD (intrauterine device), implants and sterilization [23]. Therefore, women need knowledge about all types of contraception are important, especially long-term contraception in contributing to unwanted pregnancies.

Conclusion: Based on the analysis above, the conclusions obtained are:

1. A more accurate model is the logistic regression model with SMOTE. The model without imbalance handling produces higher accuracy than the logistic regression model with SMOTE, but the specificity and G-Mean are NA values. Meanwhile, the logistic regression model with SMOTE has a higher sensitivity value and can increase the specificity and G-Mean values.

2. Using logistic regression, it is known that the variables age 25-34, age ≥ 35, education ≥ junior high school, working women, women living in urban areas, single/divorced women, and the number of children > 2 have a significant effect on unintended pregnancies.
3. The risk of women aged 25-34, aged ≥ 35, having higher education, working, married, living in a village, having living children > 2, and having sound knowledge of contraception are at a higher risk of having an unintended pregnancy.

Acknowledgement: Researchers would like to thank the Sepuluh November Institute of Technology for providing research funding under the National Institutional Development Scheme in 2023.

References

[1] E. A. Kassahun *et al.*, "Factors associated with unintended pregnancy among women attending antenatal care in Maichew Town, Northern Ethiopia, 2017," *BMC Res. Notes*, vol. 12, no. 1, pp. 1–6, 2019, doi: 10.1186/s13104-019-4419-5.

[2] A. Sarder, S. M. S. Islam, Maniruzzaman, A. Talukder, and B. Ahammed, "Prevalence of unintended pregnancy and its associated factors: Evidence from six south Asian countries," *PLoS One*, vol. 16, no. 2 February, pp. 1–13, 2021, doi: 10.1371/journal.pone.0245923.

[3] Y. Moges, S. A. Worku, A. Niguse, and B. Kelkay, "Factors Associated with the Unplanned Pregnancy at Suhul General Hospital, Northern[1] Y. Moges, S. A. Worku, A. Niguse, and B. Kelkay, 'Factors Associated with the Unplanned Pregnancy at Suhul General Hospital, Northern Ethiopia, 2018,' J. Pregnancy, vol," *J. Pregnancy*, vol. 2020, 2020.

[4] A. B. Teshale and G. A. Tesema, "Magnitude and associated factors of unintended pregnancy in Ethiopia: A multilevel analysis using 2016 EDHS data," *BMC Pregnancy Childbirth*, vol. 20, no. 1, pp. 1–8, 2020, doi: 10.1186/s12884-020-03024-5.

[5] C. Calvert *et al.*, "Risk factors for unplanned pregnancy among young women in Tanzania," *J. Fam. Plan. Reprod. Heal. Care*, vol. 39, no. 4, pp. 1–12, 2013, doi: 10.1136/jfprhc-2012-100389.

[6] F. Ebrahimzadeh, E. Hajizadeh, N. Vahabi, M. Almasian, and K. Bakhteyar, "Prediction of unwanted pregnancies using logistic regression, probit regression and discriminant analysis," *Med. J. Islam. Repub. Iran*, vol. 29, no. 1, pp. 828–832, 2015.

[7] K. Getu Melese, M. H. Gebrie, M. Berta Badi, and W. Fekadu Mersha, "Unintended Pregnancy in Ethiopia: Community Based Cross-Sectional Study," *Obstet. Gynecol. Int.*, vol. 2016, 2016, doi: 10.1155/2016/4374791.

[8] J. L. Morris and H. Rushwan, "Adolescent sexual and reproductive health: The global challenges," *Int. J. Gynecol. Obstet.*, vol. 131, pp. S40–S42, 2015, doi: 10.1016/j.ijgo.2015.02.006.

[9] N. Luthfina, "Analysis of Pregnancy Intension Risk Factors in Indonesia With Demographic Health Survey (Dhs) in 2017," *J. Biometrika dan Kependud.*, vol. 10, no. 1, p. 35, 2021, doi: 10.20473/jbk.v10i1.2021.35-44.

[10] E. Y. Boateng and D. A. Abaye, "A Review of the Logistic Regression Model with Emphasis on Medical Research," *J. Data Anal. Inf. Process.*, vol. 07, no. 04, pp. 190–207, 2019, doi: 10.4236/jdaip.2019.74012.

[11] K. Fithriasari, I. Hariastuti, and K. S. Wening, "Handling Imbalance Data in Classification Model with Nominal Predictors," *Int. J. Comput. Sci. Appl. Math.*, vol. 6, no. 1, p. 33, 2020, doi: 10.12962/j24775401.v6i1.6643.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Deep synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[13] A. Hazwani, A. Rahim, N. A. Rashid, A. Nayan, and A. Ahmad, *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*. Springer Singapore, 2019.

[14] Muthmainnah *et al.*, "What factors that cause the high numbers of unwanted pregnancy in Indonesia?," *Syst. Rev. Pharm.*, vol. 11, no. 11, pp. 1666–1671, 2020, doi: 10.31838/srp.2020.11.233.

[15] J. Merga, D. Wirtu, T. T. Bekuma, and M. T. Regasa, "Unintended pregnancy and the factors among currently pregnant married youths in Western Oromia, Ethiopia: A mixed method," *PLoS One*, vol. 16, no. 11 November, pp. 1–18, 2021, doi: 10.1371/journal.pone.0259262.

[16] S. T. Huynh *et al.*, "Prevalence of and factors associated with unplanned pregnancy among women in Koshu, Japan: Cross-sectional evidence from Project Koshu, 2011-2016," *BMC Pregnancy Childbirth*, vol. 20, no. 1, pp. 1–10, 2020, doi: 10.1186/s12884-020-03088-3.

[17] D. Belay *et al.*, "Unintended pregnancy and associated factors among unmarried female students: A case of Bahir Dar University," *Heliyon*, vol. 6, no. 6, p. e04309, 2020, doi: 10.1016/j.heliyon.2020.e04309.

[18] R. D. Wulandari and A. D. Laksono, "Faktor-Faktor yang Memengaruhi Kehamilan Tidak Diinginkan di Indonesia," *Bul. Penelit. Kesehat.*, vol. 49, no. 3, pp. 155–166, 2021, doi: 10.22435/bpk.v49i3.4650.

[19] U. Kalsum, R. Listiawaty, and D. N. A. M. Awal, "Reasons for unwanted pregnancy among women of childbearing age (15-19 years) in Jambi Province," *Kesmas*, vol. 16, no. 3, pp. 178–188, 2021, doi: 10.21109/kesmas.v16i3.4771.

[20] P. Dixit, F. Ram, and L. K. Dwivedi, "Determinants of unwanted pregnancies in India using matched case-control designs," *BMC Pregnancy Childbirth*, vol. 12, 2012, doi: 10.1186/1471-2393-12-84.

[21] B. O. Ahinkorah, "Individual and contextual factors associated with mistimed and unwanted

pregnancies among adolescent girls and young women in selected high fertility countries in sub- Saharan Africa: A multilevel mixed effects analysis," *PLoS One*, vol. 15, no. 10, pp. 1–15, 2020, doi: 10.1371/journal.pone.0241050.

[22] M. A. Habib *et al.*, "Prevalence and determinants of unintended pregnancies amongst women attending antenatal clinics in Pakistan," *BMC Pregnancy Childbirth*, vol. 17, no. 1, pp. 1–10, 2017, doi: 10.1186/s12884-017-1339-z.

[23] SDKI, "Indonesia District Health Survey 2017," p. 588, 2018.