Free Access to the whole issue

E3S Web of Conferences

Volume 475 (2024)

InCASST 2023 - The 1st International Conference on Applied Sciences and Smart Technologies

Yogyakarta, Indonesia, October 18-19, 2023

D. Widjaja and L.K. Budiasih (Eds.)

PREFACE

The 1st International Conference on Applied Sciences and Smart Technologies (InCASST 2023) has been organized by the Faculty of Science and Technology, Sanata Dharma University, Yogyakarta, Indonesia. This event was held on October 18–19, 2023, in Yogyakarta, Indonesia. As an effort to contribute in distributing research outcomes, especially in the search for renewable and clean energy, waste management, environmental management, and sustainable agriculture. InCASST 2023 presented four honorable international keynote speakers from representative countries: 1) Prof. Tokuro Matsuo, Advanced Institute of Industrial Technology - Japan; 2) Prof. Ir. Sudi Mungkasi, Ph.D., Sanata Dharma University-Indonesia; 3) Assoc. Prof. Dr. Peerapong Uthansakul, Suranaree University of Technology – Thailand, and 4) Assist. Prof. Dr. Eng. Rando Tungga Dewa, The Republic of Indonesia Defense University-Indonesia. This event selected local researchers and overseas fellows to share their best research works at this conference to reach a broader network of researchers. After a rigorous selection process, the Scientific & Editorial Board decided to publish 46 papers in E3S Web of Conferences, open-access proceedings in environment, energy, and earth sciences, managed by EDP Sciences, and indexed on Scopus, Scimago.

The published papers have passed all necessary improvement requirements following the Web of Conferences standard, reviewer's comments, and similarity tests by the Turnitin program. We want to thank the official committee, scientific & editorial boards, and organizing partners. Thanks to our co-host partners, Universitas Katolik Widya Mandala Surabaya, Universitas Prasetya Mulya, and Institut Teknologi Nasional Yogyakarta, for trusting and supporting this conference. Finally, we would like to briefly thank all presenters and attendees for their participation in sharing wonderful ideas and making creative decisions to inspire further research and exchange scientific reasons. We hope this time, all papers can be compiled into scientific works as the first publication of the 2023 InCASST. Lastly, we hope this conference encourages further research collaboration and see you at the next conference.

Assoc. Prof. Dr. Ir. Bernadeta Wuri Harini
Conference chair

InCASST

International Conference on Applied Sciences and Smart Technologies

Science and Technology Disruption in the Post Pandemic Era with Sustainable Development for Better Life Quality

18th October, 2023

Keynote Speakers

Prof.Ir. Sudi Mungkasi,Ph.D.
Sanata Dharma University
Indonesia

Assoc. Prof. Dr. Peerapong Uthansakul
Suranaree University of Technology
Thailand

Prof. Tokuro Matsuo
Advance Institute of Industrial Technology
Japan

Asst. Prof. Dr. Rando Tungga Dewa
The Republic of Indonesia Defense University
Indonesia

Call For Paper

Selected papers will be publication to E3S Web of Conferences (Scopus Indexed).

Scopes

- Renewable energy technologies and systems.
- Climate change and global warming.
- Sustainable agriculture and land use practices.
- Environmental impact assessment and management.
- Energy policy and planning.
- Clean energy and green technologies.
- Waste management and recycling.
- Air and water pollution control.
- Environmental biotechnology and microbiology.
- Carbon capture and sequestration.

Important Dates

- Extended New Submission : August 15, 2023 (Full Paper only)
- Accepted Notification : August 22, 2023
- Early Bird payment : April 30, 2023
- Late payment : August 31, 2023
- Full Paper Submission of Accepted Abstract : September 18, 2023
- Conference Day : October 18, 2023

Registration Fee :

Presenter :

- Offline :
  - Early bird : Rp. 3.250.000 / US$235
  - Late : Rp. 3.500.000 / US$250
- Online :
  - Early bird : Rp. 2.950.000 / US$185
  - Late : Rp. 3.200.000 / US$200

Non-Presenter :

- Full Board : Rp. 750.000 / US$60
- Gala Dinner : Rp. 250.000 / US$25

Payment Method :

- Paypal : suryagovinda@yahoo.com
- Bank Transfer
  - Acc. No : 0373733749
  - Acc. Name : I Gusti Ketut Puja
  - Bank Name : BCA Cabang Yogyakarta or Bank Central Asia Indonesia Jakarta 10310, Indonesia BCA Swift Code : CENAIDJA

Registration :

https://e-conf.usd.ac.id/index.php/incasst/incasst2023

Contact Persons :
Sri Hartati Wijono (+62)8112646471
Rosalia Arum Kumalasanti (+62)88232542165

Co-Host :

BINUS UNIVERSITY

Publisher :

E3S Web of Conferences

Venue :
Eastparc Hotel, Yogyakarta, Indonesia

E3S Web of Conferences **475**, 02015 (2024) https://doi.org/10.1051/e3sconf/202447502015
InCASST 2023

Classification of delivery type of pregnant women using support vector machine

Maria Yubela Chelsea[1*], and *Paulina H. Prima* Rosa[1]

[1] Informatics Department, Sanata Dharma University, Paingan Maguwoharjo Depok Sleman Yogyakarta, Indonesia

Abstract. One of the ways to reduce maternal mortality is by diagnosing childbirth to find out whether a mother will give birth normally or not so that appropriate treatment can be done. This study aims to improve maternal safety and health by classifying delivery type of pregnant women, either Caesarean or normal types, using the Support Vector Machine method. The dataset used in this study was taken from a hospital in 2020. It consists of 25 attributes and 302 records that include information about the health conditions of pregnant women and babies. Several experiments were performed towards the dataset with and without balancing. Three types of SVM kernels, namely Linear, RBF, and Polynomial kernels, were then implemented to classify the dataset using several variations of parameters of C, gamma, and degree. The validation was performed using several k-fold cross validations. The results of this study show that the highest accuracy is 92.98% at 5-fold cross validation using the RBF kernel with parameters C = 10 and gamma = 1. The performances of the three SVM kernels varied depending on the type of data used.

1 Introduction

Pregnant women can give birth through one of the following delivery types, namely normally or through cesarean section. Giving birth normally means giving birth on time, without any administration of certain drugs. While cesarean section is the process of removing the baby from the uterus without going through the birth canal or vagina, but through surgery performed on the abdomenand uterine wall [1]. If the type of delivery is not predicted in advance, it may have a fatal impact on the safety of both mother and baby. Therefore, a diagnosis of childbirthis necessary to know whether a mother will give birth normally or not. The results of the diagnosis can be usedas a basis for decisions to take further action. Thus, handling can be quickly and precisely carried out in accordance with the decision of action.

Based on the above problems, technology might playa role. In information technology, there are classification techniques that can help to classify the type of labor. One method that can be used is Support Vector Machine (SVM). SVM is a technique to find a hyperplane or barrier that can separate two data sets from two different classes. SVM has advantages including determining distances using support vectors so that the computational process becomes fast [2].

* Corresponding author: chelseayubela@putranirmala.sch.id

This study aims to improve maternal safety and health by classifying delivery type of pregnant women, either Caesarean or normal types, using the Support Vector Machine method. Testing was carried out using

4 variations of datasets, namely raw data without balancing, raw data with balancing, selected data without balancing, and selected data with balancing. The model was experimented using 3 kernels in SVM, namely linear kernels, RBF, polynomials and using several variations of parameters of C, gamma, and degree. The validation was performed using several k- fold cross validations. The best model will then can be used to classify new data records.

2 Support vector machine

Support Vector Machine (SVM) can be used for linear and non-linear data separation by utilizing kernel techniques [3]. In this study, this dataset is classified asnon-linear problem, because it has many attributes and has unique and complex relationships. Therefore, in this study, the classification of delivery types of pregnant women will utilize the function of the kernel technique.The 3 kernels used are linear, RBF, and polynomial.

Linear kernel functions generate a linear hyperplaneinside an extended feature space. A linear kernel is the simplest kernel function. The linear kernel is used whenthe data being analyzed can be separated by a straight line [3].

Polynomial kernel functions project data into extended feature spaces using polynomial functions.Polynomial kernel functions are used when data cannotbe separated linearly. This kernel function measures the similarity betweentraining sample vectors in a feature space. Polynomial kernels are also suitable for solving classification problems on normalized training datasets [4].

Gaussiankernel functions, also known as Radial Basis Functions(RBFs), project data into an infinite feature space usingGaussian functions. RBF kernels are one of the most common types of kernels used to solve classification problems on data that cannot be separated linearly.

3 Methodology

3.1 Dataset

The dataset used in this study was data on pregnant women of a hospital in North Bengkulu. The data has 25 attributes, namely age, gestational age, leukocytes, lymphocytes, hemoglobin, platelets, erythrocytes, blood pressure, glucose at times, ureum, SGOT, SGPT, urine glucose, protein, single/double, baby position, narrow pelvis, premature rupture of membranes, umbilical cord, asthma, hepatitis, HIS (contractions), history of partus, and amniotic condition.

The dataset used in this study was imbalanced, since the number of caesarean cases are 260, while the normal cases are 42.Therefore, the data is subjected to a balancing process. The goal of this process is to raise the minorityclass to have a balanced number with the majority class, so that SVM models will not tend to train data from the majority class [5]. The dataset will also be selected using information gain. Several experiments were implemented using different dataset as explained further in the following section.

 https://doi.org/10.1051/e3sconf/202447502015

3.2 Research design

Figure 1 shows the design of this research in general. Raw data undergoes several pre-processes, namely data cleaning, data transformation, and data selection. Those processes resulted on four type of dataset. Dataset 1 is raw data that undergoes cleaning and transformation only. Dataset 2 undergoes balancing process after cleaning and transformation. Dataset 3 undergoes selection attributes after cleaning and transformation, but does not undergo balancing process. While dataset 4 undergoes cleaning, transformation, selection attributes, and balancing. To observe the impacts of cleaning, transformation attributes selection, and balancing towards accuracy, classification process using SVM are performed towards all dataset by utilizing linear kernels, RBF, polynomials and using several variations of parameters of C, gamma, and degree. The accuracy of each experiment are compared and analyzed to find the best one.

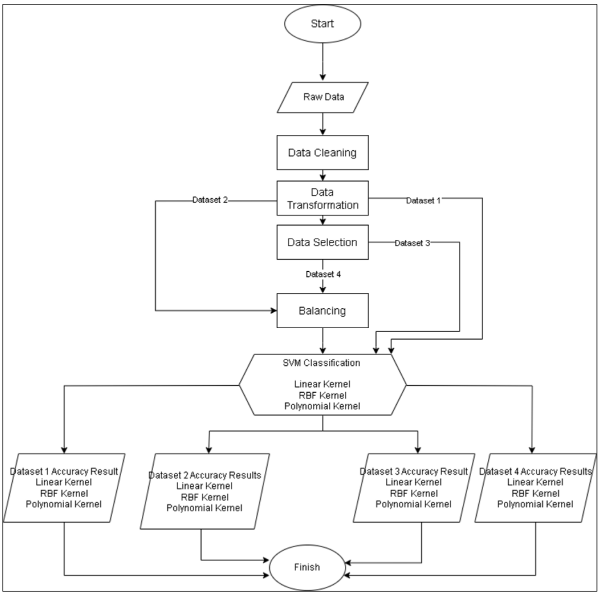

Fig. 1. Research design

3.3 Data preprocessing

3.3.1 Data cleaning

Data was cleaned to avoid noise or missing values which can interfere with the classification process. The data was checked using the isna function in Python. If the data was detected having noise, it was cleaned using the filna function in Python. Empty data was replaced by most frequent values of the related variable.

3.3.2 Data transformation

The data was transformed into numeric. The transformation was carried out based on certain intervals from previous studies [6], halodoc.com, and laboratory test results of the hospital. Table 2 describes the transformed attributes.

3.3.3 Data selection

At this stage, 25 attributes were ranked with the information gain function in python. Attributes that have an information gain value of more than 0 were then selected.

3.3.4 Data balancing

A balancing process was performed since the dataset in this study was imbalanced. Fig. 2 and Fig. 3 show a comparison of classes before and after balancing.

Table 2. Description of transformed attributes.

No	Atrribute	Description
1	Age	20 - 35 years = 2 If age < 20 and > 35 = 1
2	Gestational age	37 - 42 weeks = 2 If gestational age < 37 and >42 = 1
3	Leukocytes (WBC)	6000 - 13000 microliters = 2 Ifleukocytes (WBC) < 6000 and > 13000 = 1
4	Lymphocytes (LYM)	15 - 50 military = 2 If lymphocytes(LYM) < 15and > 50 = 1
5	Haemoglobin (HGB)	11.5 - 16.5 Gram/dL If hemoglobin (HGB) < 11.5and > 16.5 = 1
6	Thrombocytest (PLT)	130 - 400 milliliters If platelets (PLT) < 130 and >400 = 1
7	Erythrocytes (RBC)	3.5 - 5.5 cells/mm3 If Erythrocytes (RBC) < 3.5and > 5.5 = 1
8	Blood Pressure	Normal = 1At risk = 2
9	Intermittent glucose	≤ 200 mg/dL = 2 ≥ 200 = 1
10	Ureum	18 - 48 md/dL = 2 If ureum < 18 and > 48 = 1
11	SGOT	5-40 micro/liter = 2 If SGOT < 5 and > 40 = 1
12	SGPT	7-57 micro/liter = 2 If SGPT < 7 and > 57 = 1
13	Urine glucose	Negative = 1, Positive = 2
14	Protein	Positive = 1, Negative = 2
15	Single/Double	JTH (Single Live Fetus) = 1Gameli = 2
16	Baby position	Preskep = 1, Breech = 2
17	Narrow pelvis(CPD)	No = 1, Yes = 2
18	Early rupture of membranes	No = 1, Yes = 2
19	Umbilical cord	Normal = 1, Entangled = 2
20	Asthma	Negative = 1, Positive = 2
21	Hepatitis (HbaSG)	Negative = 1, Positive = 2
22	HIS (Contraction)	Negative = 1, Positive = 2
23	Partus history	Caeserean = 1, Spontaneous = 2, None = 3
24	Amniotic condition	Normal = 1, Few = 2, Many =3
25	Partus	Spontaneous = 1, SC = 2

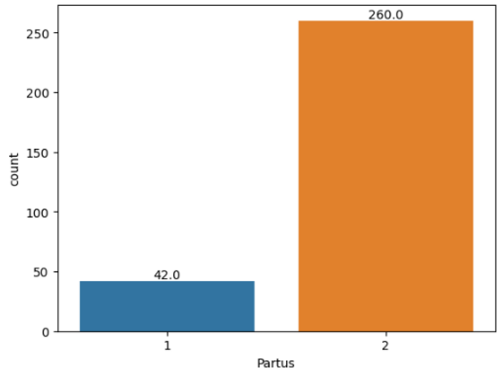

Fig. 2. Dataset Before Balancing

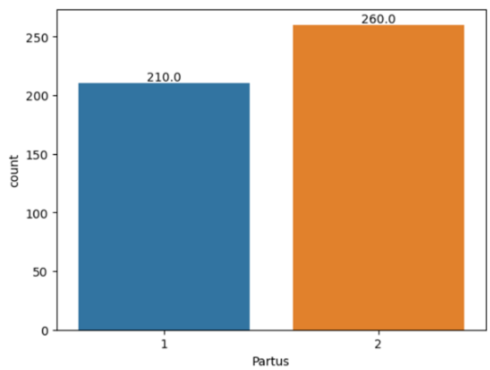

Fig. 3. Dataset After Balancing

3.3.5 Classification using SVM

The classification using SVM was performed in the following steps:

1) Initialization of several variables containing a list of k-fold values, C values, gamma values, and degree values to be tested in the experiment.
2) Nested loops were performed to run experiments on various combinations of k-fold, C, gamma, and degree values. At each iteration, SVM testing was performed with each kernel and corresponding parameter values, using k-fold cross-validation. After that, the classification accuracy and confusion matrix of each k-fold iteration were recorded in a list.
3) After the loop was complete, the average accuracy and average confusion matrix for each kernel were calculated

4 Results and analysis

The analysis were performed towards three point of views:

1) finding the highest accuracy of each dataset
2) comparing the performance of algorithms towards dataset with and without balancing
3) comparing the performance of the three kernels.

4.1 Finding the highest accuracy of each dataset

Table 3 describes the summary of highest accuracy for each dataset. Raw data with balancing results on the best accuracy among the four datasets.

Table 3. Summary of the Highest Accuracy of Each Dataset

Data	Kernels	Accuracy	Number of K-Folds
Raw datawithout balancing	Linear (C=10)	90.06%	10
Raw datawith balancing	RBF (C=1.0, *Gamma*=1)	94.26%	5
Selecteddata without balancing	Polynomial (C=10, *Degree* =2)	91.06%	10
Selected data with balancing	RBF (C=10, *Gamma*=1)RBF (C=100, *Gamma*=1)	91.06%	5

4.2 Performance comparison of dataset with and without balancing

Figure 4 shows that for raw dataset, balancing gives impacts on the improvement of accuracy compare to data without balancing. While for selected data, there is no difference between data with and without balancing.

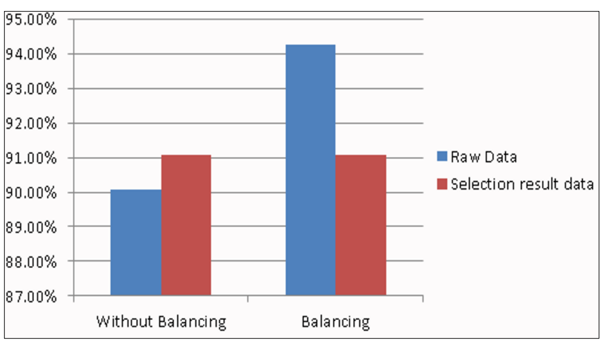

Fig. 4. Performance Comparison of Dataset With and Without Balancing

4.3 Performance comparison of three kernels

Figure 5 shows that the performance of 3 kernels varies depending on the type of data. The Linear Kernel excels at data without balancing. The RBF kernel excels at balancing data. Polynomial kernels show competitive performance on both types of data, especially on selected data in balancing.

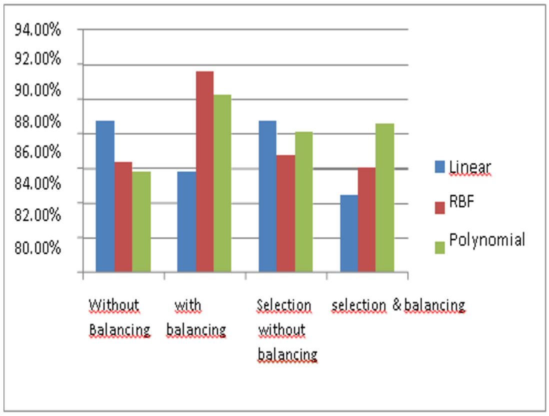

Fig. 5. Performance Comparison of Three Kernels

Fig.5. describes the differences in performance among commonly used kernels in Support Vector Machine (SVM) algorithms for data processing. The Linear Kernel is the simplest and excels in handling imbalanced data, particularly when there's a significant number of samples in each class, as it efficiently separates linearly separable classes. The RBF (Radial Basis Function) Kernel is suitable for dealing with imbalanced data and complex separation patterns, adapting well to non-linearly distributed data. The Polynomial Kernel offers adaptable complexity and competitive performance on both balanced and imbalanced data by adjusting the polynomial degree to suit the dataset. When selecting an SVM kernel, it's essential to understand the data's characteristics. When dealing with imbalanced data, choosing the RBF kernel or adjusting the polynomial kernel wisely may be a better option, but there are no strict rules, and experimenting with different kernels is a good approach to determine the most suitable one for each specific case.

Based on the results of this study, the best accuracy of the experiment in this research, which is 94.26%, shows that SVM has a better ability to classify the dataset into two classes, when compared to the decision tree algorithm in the research conducted by Gharehchopogh et.al. whose success rate is 86.25% [7] and the research by Kamat et.al. whose precision is 92.9% [8].

The best accuracy of the experiment in this research, is not far from the accuracy of research performed by Amin and Ali [9] that evaluated the performance of random forest, logistic regression, naive bayes, k-nearest neighbors, and support vector machine methods on cesarean data using 80 data. In their research, the highest success rate was found in 95% k-nearest neighbors and random forest.

In addition, this research has added more attributes and more number of records compared to the research performed by previous researches. In this study, the authors found that implementing attribute selection using information gain did not improve the accuracy. Therefore, 25 attributes were used to build the model. These attributes are age, gestational age, leukocytes, lymphocytes, hemoglobin, platelets, erythrocytes, blood pressure, glucose during pregnancy, ureum, SGOT, SGPT, urine glucose, protein, single/double, baby position, narrow pelvis, premature rupture of membranes, umbilical cord, asthma, hepatitis, HIS (contractions), history of parturition, premature rupture of membranes, and partus.

Whereas the research of Çelik [10] showed that by implementing Artificial Neural Network using 56 data, the accuracy of training dataset was 96.43% and the accuracy of testing data was 62.5%. Çelik also concluded that in performing a caesarean section, the features with the highest level of effect were determined as the number of births and duration of birth, the least affecting features were age and blood pressure, and the moderate effect was heart problems. It becomes a challenge for the future research to improve the performance, especially towards testing data.

5 Conclusion

In conclusion, the study demonstrates the effectiveness of Support Vector Machine (SVM) in classifying the type of delivery in pregnant women, achieving the highest accuracy of 94.26% using the RBF kernel with parameters C = 1.0 and gamma = 1 in a 5-fold cross-validation setting. The performance of three different kernels, namely Linear, RBF, and Polynomial, varies according to the characteristics of the data. The Linear Kernel performs well on imbalanced data, while the RBF kernel excels in handling balanced data, and Polynomial kernels offer competitive results on both data types, particularly on selected balanced data. The study highlights the significant impact of data balancing on classification accuracy. Surprisingly, implementing feature selection based on information gain did not lead to improvement in accuracy, emphasizing the pivotal role of kernel selection and data balancing in SVM-based classification tasks. Future research might be performed to improve the accuracy of research using independent data testing and concentrated on more specific attributes according to Çelik.

References

1. A. Kurniarum, Asuhan Kebidanan Persalinan dan Bayi Baru Lahir, 1 st ed., 3 - 8 , (2 0 1 6)
2. C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2, 121–167, (1998), accessed from https://doi.org/10.1023/A:1009715923555
3. B.Raharjo, Pembelajaran Mesin, In Angewandte Chemie International Edition. 951–952, (2016)
4. R.D. Mukarramah, D . Atmajaya, L. B. Ilmawan. Perfomance Comparison of Support VectorMachine (SVM) with Linear Kernel and Polynomial Kernel for Multiclass Sentiment Analysison Twitter. ILKOM Jurnal Ilmiah 13 (2), 168-174, (2021), accesed from https://doi.org/10.33096/ilkom.v13i2.851. 168-174
5. R. F. W. Pratama, Boosting Support Vector Machine pada Data Microarray yang Imbalance.PhD Thesis Institut Teknologi Sepuluh Nopember, (2018)
6. A. Armonica A, P.H.P. Rosa, Klasifikasi Jenis Persalinan pada Ibu Hamil dengan Metode Random Forest, Prosiding Seminar Nasional Teknologi Informasi dan Bisnis. Surakarta. (2022)
7. F.S. Gharehchopogh, Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study. Int. J. of Computer Applications, 52-No.6, 0975 – 8887, August, (2012).
8. A. Kamat, V. Oswal, M. Datar, Implementation of Classification Algorithms to Predict Mode of Delivery, Int. J. of Computer Science and Information Technologies, 6 (5), 4531-4534, (2015).
9. M.Z. Amin, A. Ali, Performance Evaluation of Supervised Machine Learning Classifier for Predicting Healthcae Operational Decisions, C-Section Classification Database Report, UCI Machine Learning Repository, University of California, Irvine, USA, Technical Report, Januari (2017).

10. S. Çelik, Determination and Classification of Importance of Attributes Used in Diagnosing Pregnant Women's Birth Method. Alphanumericjournal, 8, Issue 2, (2020).