Predicting Awareness and Misconceptions on Sexual Education among Pre-Service Teachers Using Naive Bayes Classification

Kyza L. Quimpan, Romulo L. Lanugan

Southern Leyte State University-Tomas Oppus Campus, Bontoc, Southern Leyte, Philippines

ABSTRACT

This study seeks to fill that gap by evaluating the awareness and misconceptions of student teachers regarding sexual education. This employed a quantitative descriptive research design integrated with a machine learning approach to examine the awareness and misconceptions of pre-service teachers regarding sexual education. A Naive Bayes classifier—specifically the BayesNet implementation in Weka with the K2 search strategy and Simple Estimator—was used to predict awareness levels based on participant responses and demographic profiles. The Naive Bayes classifier achieved a strong overall performance in predicting the awareness levels of pre-service teachers on sexual education, These findings have direct implications for teacher education institutions. By integrating predictive analytics into assessment processes, programs can accurately map awareness levels among their students, enabling more strategic allocation of training resources.

Keywords: sexual education, teacher education, pre-service teacher attitudes, Naive Bayes classifier, predictive modelling,

INTRODUCTION

Sexual education is defined as an age-appropriate, culturally relevant approach to teaching and learning about sex and relationships by providing scientifically accurate, realistic, and non-judgmental information. It goes beyond basic biological concepts such as reproduction or disease prevention and aims to promote a holistic understanding of human sexuality, relationships, and personal responsibility. According to the World Health Organization (2023), *"Comprehensive sexuality education (CSE) gives young people accurate, age-appropriate information about sexuality and their sexual and reproductive health, which is critical for their health and survival."*

Student teachers, like any beneficiaries of education systems, have the right to access comprehensive sexuality education—especially during their formative post-puberty years. This education benefits not only their personal sexual health but also prepares them to be effective sexuality educators. According to the study of Bruno et al (2024), review results provide evidence of teachers' challenges in implementing sexuality education (SE) programs, teachers' opinions on SE, and the importance of including all educational figures in the SE programs.

Despite its importance, sex education remains a sensitive and underdeveloped topic in many regions, especially in Asian and conservative contexts. Cultural taboos, religious objections, and a lack of trained educators hinder effective delivery. Costello et al (2022) & O'Brien et al (2020) agreed that training for SE is limited and SE provision received is varied and not reflective of comprehensive SE. To add, teachers often report discomfort and insufficient support when tasked with discussing sexual health topics, resulting in low-quality implementation (Leung et al., 2019). Additionally, debates continue about which educational model to adopt—be it abstinence-only, abstinence-until-marriage, or comprehensive sex education (CSE). While CSE

promotes gender equality, civic engagement, and rights-based learning, many programs still fall short of addressing the actual needs and awareness levels of students.

Although several studies have investigated sexual education, many have overlooked its perception and awareness among pre-service teachers, who will one day serve as facilitators of such instruction. This study seeks to fill that gap by evaluating the awareness and misconceptions of student teachers regarding sexual education. To analyze the patterns in their responses, the Naive Bayes classifier was applied, a probabilistic machine learning model well-suited for categorical survey data. This method enables to predict levels of awareness based on various demographic and attitudinal variables, offering data-driven insights into existing knowledge gaps. The findings can inform the development of more effective, targeted sex education programs in teacher training institutions.

LITERATURE REVIEW

Sexual education is widely acknowledged as a vital component of a holistic and effective educational system. It goes beyond the basic transmission of biological and reproductive knowledge and aims to equip young people with the understanding, attitudes, and skills necessary to make informed decisions about their sexual health and relationships. According to the World Health Organization (2023), sexual health is not merely the absence of disease or dysfunction but the presence of physical, emotional, mental, and social well-being in relation to sexuality. Student teachers, as future facilitators of sexual education, are especially important to consider in this context. They are not only recipients of sexuality knowledge but also positioned to transmit that knowledge to future generations. Hence, their awareness, attitudes, and misconceptions must be thoroughly examined.

Several studies have highlighted how educators' personal beliefs and attitudes shape how they deliver sexuality education (SE). In the study of Costello et al. (2022), results revealed that research on SE during Initial Teacher Education (ITE) is limited and minimal research has focused on student teachers' attitudes on SE. In a study by Huertas-Abril & Palacios-Hidalgo (2022), shows that even when pre-service teachers demonstrate positive views regarding the LGBTIQ+ community, their lack of awareness or experience with these issues may hinder their teaching practice.This distinction is particularly important for student teachers who may still be in the process of forming their own understanding of sexual health and morality. The present study builds on these insights by examining pre-service teachers' awareness levels using a data-driven approach, enabling the identification of factors that correlate with either accurate understanding or persistent misconceptions.

Cultural and religious barriers often hinder the effective implementation of sexual education, particularly in Asian and conservative contexts. The idea of sex as a taboo topic, often rooted in religious or traditional beliefs, creates discomfort among teachers and resistance among communities. Many still believe that discussing sexuality encourages promiscuity or deviates from moral teachings. As a result, teachers may feel unequipped or even fearful of addressing such topics in the classroom. Escrig-Estrems et al. (2025) argues that SE training is necessary for pre-service teachers to be prepared to teach sensitive topics. Further, Leung et al. (2019) found that educators often cite religious conflict, lack of training, and insufficient institutional support as major barriers to delivering sexual health education effectively. These systemic issues contribute to the persistence of misconceptions and the inadequate implementation of comprehensive sex education programs.

Furthermore, various misconceptions about sexual education continue to circulate, particularly in regions with limited access to accurate, values-based instruction. For example, beliefs that sexual education encourages promiscuity, corrupts youth morality, or undermines abstinence are still prevalent. These assumptions often shape not only public opinion but also policy decisions and curriculum design. This study aims to uncover the extent of these misconceptions among student teachers by applying a Naive Bayes classifier—a machine learning model capable of identifying patterns in categorical data such as survey responses. While traditional survey analysis provides descriptive insights, Naive Bayes allows for a more predictive analysis, offering a new perspective on which variables are most associated with accurate awareness or misunderstanding.
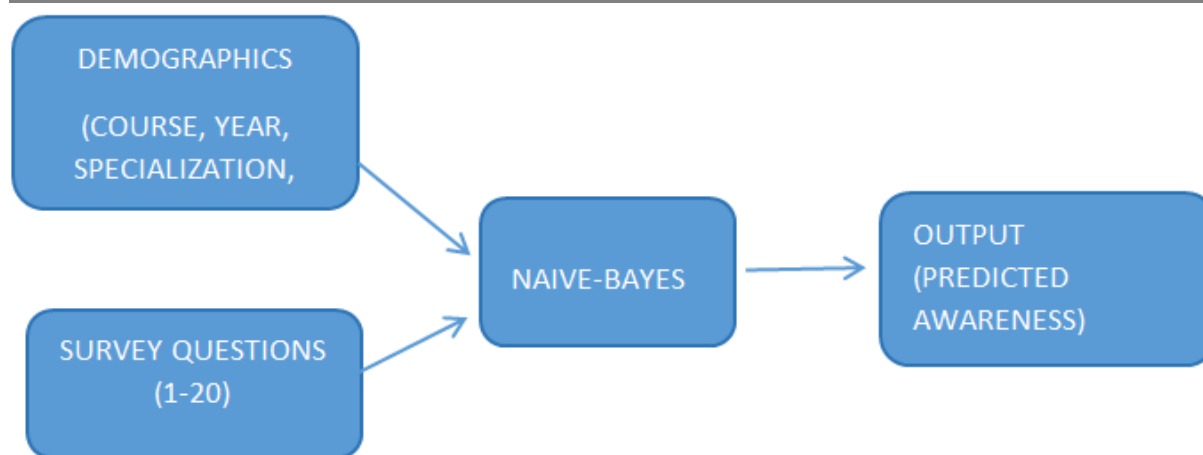
Fig. 1. Paradigm of the Study

METHODOLOGY

This study employed a quantitative descriptive research design integrated with a machine learning approach to examine the awareness and misconceptions of pre-service teachers regarding sexual education. A Naive Bayes classifier—specifically the BayesNet implementation in Weka with the K2 search strategy and Simple Estimator—was used to predict awareness levels based on participant responses and demographic profiles.

Naive Bayes is a probabilistic classification algorithm based on Bayes' Theorem, which calculates the probability that a given instance belongs to a specific category. It assumes that all features (e.g., survey responses, demographic variables) are conditionally independent given the class label—an assumption that simplifies computation and often works well even when it is not perfectly true.

In this study, the BayesNet implementation was used, which extends the basic Naive Bayes approach by representing the relationships between variables as a Bayesian network. Instead of assuming total independence between features, BayesNet can model some dependencies, potentially improving classification accuracy. The K2 search strategy was applied to determine the optimal structure of the network by adding edges between related variables, and the Simple Estimator was used to calculate the conditional probability tables (CPTs) for each node in the network.

Participants consisted of 129 pre-service teachers enrolled in a teacher education institution, selected through purposive sampling to ensure representation across various specializations and year levels. Data were collected via a structured questionnaire composed of two sections:

1. Demographic information – course, specialization/major, year level, and sexual orientation.

2. Awareness and belief statements – 20 closed-ended statements assessing knowledge and misconceptions about sexual education, rated on a Likert scale.

The instrument underwent expert validation for content accuracy and clarity. Informed consent was obtained, and participant anonymity was maintained throughout the study.

Data Preparation and Analysis

Responses were encoded into nominal form, and missing or inconsistent entries were removed. The final dataset contained 24 attributes (four demographic variables and 20 survey items) and 129 instances. The model was trained and tested using 10-fold cross-validation in Weka to ensure reliable performance estimates.

Performance was evaluated using classification accuracy, Kappa statistic, precision, recall, F-measure, ROC area, and confusion matrix. This combination of metrics provided a comprehensive view of the model's ability to correctly classify awareness levels and to handle imbalanced or overlapping categories.

Given the sensitivity of sexual education as a research topic, strict ethical guidelines were observed throughout the conduct of this study. Informed consent was obtained from all participants after providing them with a clear explanation of the study's objectives, scope, and procedures, while ensuring that participation was voluntary and that they retained the right to withdraw at any time. To preserve anonymity and confidentiality, all personal identifiers were removed from the dataset and responses were coded, with the data securely stored and accessible only to the researchers. The questionnaire was designed to be neutral, respectful, and culturally sensitive so as not to reinforce stigma, bias, or discomfort among respondents. Participants were also reminded that they could decline to answer any item they found intrusive, and counseling referrals were made available to mitigate potential psychological distress. The ethical use of data analytics was also emphasized, as the Naive Bayes model was employed strictly for academic purposes, with findings presented only in aggregate form to prevent any individual from being singled out or stigmatized based on their awareness level. Furthermore, the study underwent expert validation and adhered to the institutional research ethics policies prior to data collection, ensuring that the rights, dignity, and well-being of participants were safeguarded at all stages of the research process.

RESULT AND DISCUSSION

The Naive Bayes classifier was applied to the dataset consisting of 129 instances and 24 attributes, including demographic variables (Course, Specialization/Major, Year/Level, Sexual Orientation) and 20 survey statements (S1–S20). The classification task was evaluated using 10-fold cross-validation. The results are summarized below.

Summary of Classifier Performance

Metric	Value
Correctly Classified Instances	108 (83.72%)
Incorrectly Classified Instances	21 (16.28%)
Kappa Statistic	0.7878
Mean Absolute Error	0.0605
Root Mean Squared Error	0.2266
Relative Absolute Error (%)	19.585%
Root Relative Squared Error (%)	57.6845%
Total Instances	129

The model achieved a classification accuracy of 83.72%, correctly predicting 108 out of 129 instances. The Kappa statistic of 0.7878 indicates substantial agreement between predicted and actual classes beyond chance. The low Mean Absolute Error (0.0605) and Root Mean Squared Error (0.2266) demonstrate high prediction reliability, with relative errors indicating only small deviations from the true values. Overall, the model exhibits strong predictive capability for classifying awareness levels among pre-service teachers.

Class-wise performance metrics of the Naive Bayes classifier

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area
A (Aware)	0.879	0.031	0.906	0.879	0.892	0.992	0.980
FNA (Fairly Not Aware)	0.783	0.000	1.000	0.783	0.878	0.994	0.977
NeA (Not Enough Awareness)	0.692	0.029	0.857	0.692	0.766	0.962	0.914
NA (Not Aware)	0.895	0.132	0.739	0.895	0.810	0.975	0.947
FA (Fully Aware)	1.000	0.025	0.750	1.000	0.857	0.999	0.989
Weighted Average	0.837	0.054	0.853	0.837	0.837	0.982	0.957

Class-wise Detailed Accuracy

Class A (Aware) and FA (Fully Aware) achieved both high recall and high precision, meaning the model can accurately identify individuals with high awareness while avoiding false positives.

FNA (Fairly Not Aware) achieved perfect precision (1.000), indicating no false positives, though recall is slightly lower (0.783), meaning some members of this group were misclassified as other categories.

NeA (Not Enough Awareness) and NA (Not Aware) improved significantly compared to the lower-accuracy model. NA achieved high recall (0.895), showing strong detection ability, though with slightly lower precision (0.739).

All classes achieved high ROC Areas (>0.96), suggesting excellent class separability and robust classification thresholds.

Confusion Matrix

	a	b	c	d	e
a = A	29	0	0	2	2
b = FNA	1	18	0	4	0
c = NeA	1	0	18	6	1
d = NA	1	0	3	34	0
e = FA	0	0	0	0	9

The confusion matrix shows that most classifications were correct, with only minor overlaps:

NA was occasionally misclassified as NeA (3 cases) and vice versa (6 cases).

FNA was sometimes mistaken for NA (4 cases).

Class A had minimal confusion, with only 4 total misclassifications.

DISCUSSION

The Naive Bayes classifier achieved a strong overall performance in predicting the awareness levels of pre-service teachers on sexual education, with an accuracy of 83.72% and a Kappa statistic of 0.7878, indicating substantial agreement beyond chance. The low Mean Absolute Error (0.0605) and Root Mean Squared Error (0.2266) further confirm the reliability of the predictions. These results demonstrate that the model can effectively classify respondents into the five awareness categories—Fully Aware (FA), Aware (A), Fairly Not Aware (FNA), Not Enough Awareness (NeA), and Not Aware (NA)—based on demographic and survey response data.

From an educational standpoint, the ability to accurately identify awareness levels among pre-service teachers is highly valuable. Teacher training institutions can use such predictive insights to design targeted interventions, ensuring that future educators not only possess accurate knowledge about sexual education but can also deliver it confidently and without bias.

The model performed exceptionally well for FA and A categories. FA achieved perfect recall (1.000), meaning all respondents who were fully aware were correctly identified, while maintaining a strong precision score of 0.750. The A class showed both high recall (0.879) and precision (0.906), indicating that most well-informed students were classified correctly and with minimal confusion. These groups represent pre-service teachers who already demonstrate strong or comprehensive knowledge of sexual education and could potentially serve as peer mentors or role models in training activities.

The FNA group also showed promising results, achieving perfect precision (1.000), meaning the model did not incorrectly label any student as Fairly Not Aware. However, recall (0.783) suggests that some members of this category were classified elsewhere, possibly due to overlapping response patterns with higher- or lower-awareness groups.

The NeA category remains the most challenging for the model, with a recall of 0.692. This indicates that nearly one-third of students with partial awareness and significant misconceptions were misclassified, often as

NA or A. This is a critical finding, as these individuals may be overlooked in training needs assessments, leading to persistent gaps in knowledge.

The NA category had a high recall (0.895) but lower precision (0.739), meaning that while most students with low awareness were correctly detected, a notable proportion of those predicted as NA actually belonged to other categories. This suggests a need for better feature refinement to improve differentiation between NA and the mid-awareness groups (NeA, FNA).

These findings align with existing literature highlighting that awareness and understanding of sexual education among pre-service teachers are inconsistent and often influenced by cultural taboos, personal values, and lack of formal training (Leung et al., 2019). The model's high performance in detecting high-awareness individuals allows institutions to recognize and leverage these students in peer-assisted learning, while its capacity to identify low- and mid-awareness groups provides a basis for targeted curriculum interventions.

By applying predictive analytics, teacher education programs can allocate resources more strategically—intensifying foundational instruction for NA and NeA groups, providing corrective training for FNA, and engaging FA and A students as facilitators in awareness-raising initiatives. In doing so, institutions can better prepare future teachers to deliver comprehensive, accurate, and culturally sensitive sexual education in their classrooms.

Based on the findings of the study, it is recommended that teacher education institutions enhance their curriculum by integrating structured modules on comprehensive sexuality education (CSE), with particular attention to correcting misconceptions and addressing the needs of students who fall under the "Not Enough Awareness" and "Not Aware" categories. Targeted intervention programs such as remedial workshops, seminars, and peer-assisted learning activities may be introduced, wherein students who belong to the "Fully Aware" and "Aware" groups can serve as mentors or resource persons for their peers. Moreover, faculty members should be provided with continuous professional development opportunities that focus on effective strategies for teaching sensitive topics in a culturally sensitive and unbiased manner. At the policy level, the adoption of predictive assessment tools, such as the Naive Bayes classifier applied in this study, can support evidence-based decision-making by providing data-driven insights into students' awareness levels. Finally, future research may expand this study by including a larger and more diverse sample across different regions and cultural contexts, as well as by incorporating additional variables such as religiosity, previous exposure to sex education, or family background, in order to refine the predictive model and enhance the generalizability of the findings.

CONCLUSION

This study applied a Naive Bayes classifier to predict the awareness levels and misconceptions of pre-service teachers regarding sexual education, using demographic data and responses to a structured survey. The model achieved a strong classification accuracy of 83.72% with a Kappa statistic of 0.7878, indicating substantial agreement beyond chance. All classes demonstrated high ROC values (≥ 0.962), confirming the model's robust ability to distinguish between awareness categories.

The results reveal that the model is highly effective in identifying pre-service teachers with high awareness levels—specifically the Fully Aware (FA) and Aware (A) groups—while also successfully detecting those with lower awareness, such as Not Aware (NA) and Fairly Not Aware (FNA). The greatest classification challenge lay with the Not Enough Awareness (NeA) category, which showed moderate recall, highlighting a need for more targeted interventions to address partial knowledge and misconceptions.

These findings have direct implications for teacher education institutions. By integrating predictive analytics into assessment processes, programs can accurately map awareness levels among their students, enabling more strategic allocation of training resources. High-awareness individuals can be engaged as peer facilitators or resource persons, while those in lower-awareness categories can receive tailored instruction aimed at correcting misconceptions and improving content mastery.

Ultimately, this approach contributes to the preparation of future educators who are not only knowledgeable about sexual education but also capable of delivering it effectively, confidently, and sensitively. By closing awareness gaps during teacher training, institutions can help ensure that comprehensive sexuality education is implemented with both accuracy and cultural competence in school settings.

REFERENCES

1. Bruno, V., Baiocco, R., & Pistella, J. (2024). Teachers' Attitudes and Opinions Toward Sexuality Education in School: A Systematic Review of Secondary and High School Teachers. American Journal of Sexuality Education, 20(3), 390–428. https://doi.org/10.1080/15546128.2024.2353708
2. Costello, A., Maunsell, C., Cullen, C., Bourke, A. (2022). A Systematic Review of the Provision of Sexuality Education to Student Teachers in Initial Teacher Education. Sec. Teacher Education. Vol. 7 - 2022. https://doi.org/10.3389/feduc.2022.787966
3. Escrig-Estrems, A., & Talavera, M. (2025). Training in Comprehensive Sexuality Education Received by Pre-Service Teachers: A Scoping Review. American Journal of Sexuality Education, 1–18. https://doi.org/10.1080/15546128.2025.2532767
4. Huertas-Abril, C. A., & Palacios-Hidalgo, F. J. (2022). LGBTIQ+ issues in teacher education: a study of Spanish pre-service teachers' attitudes. Teachers and Teaching, 28(4), 461–474. https://doi.org/10.1080/13540602.2022.2062740
5. IBM. (2023). What are Naive Bayes classifier? https://www.ibm.com/think/topics/naive-bayes
6. Leung, H., Shek, D. T., Leung, E., & Shek, E. Y. (2019). Development of contextually-relevant sexuality education: Lessons from a comprehensive review of adolescent sexuality education across cultures. International Journal of Environmental Research and Public Health, 16(4), 621. https://doi.org/10.3390/ijerph16040621.
7. O'Brien, H., Hendriks, J., & Burns, S. (2020). Teacher training organisations and their preparation of the pre-service teacher to deliver comprehensive sexuality education in the school setting: a systematic literature review. Sex Education, 21(3), 284–303. https://doi.org/10.1080/14681811.2020.1792874
8. World Health Organization. (2023). What is comprehensive sexuality education? https://www.who.int/news-room/questions-and-answers/item/comprehensive-sexuality-education