International Journal of Engineering and Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 15 Issue 8 August 2026 || PP 09-16

Multiple Imputation Strategies for Nonignorable Missing Data Using Data Augmentation in Industrial Analytics

Ms Surabhi S, Dr T. A. Ashok Kumar, Jeevitha S
Postgraduate Studies, CMR University, Bengaluru.
Professor, CMR University, Bengaluru.
Assistant Professor, Excel College of Engineering, Salem.

Abstract
Abstract: *The presence of incomplete observations is a persistent challenge in industrial data analysis, often reducing the reliability of statistical inference, predictive modeling, and data-driven decision-making. Conventional methods for handling missing values, such as case deletion or single imputation, frequently introduce bias and result in significant information loss. Although Multiple Imputation (MI) has been widely adopted as a robust approach for estimating missing values, its performance is generally based on the assumption that data are Missing at Random (MAR). In many industrial environments, however, missingness is influenced by unobserved variables, resulting in a Missing Not at Random (MNAR) mechanism that limits the effectiveness of traditional MI techniques. This study proposes an enhanced multiple imputation framework that integrates a selection model with data augmentation and semi-parametric estimation to effectively address nonignorable missing data. The proposed framework jointly models the data generation process and the missingness mechanism, thereby improving the estimation of incomplete observations under MNAR conditions. In addition, the incorporation of the Darboux variate constrains imputed values within feasible lower and upper limits, preserving the statistical characteristics and structural integrity of the original dataset. The proposed methodology enhances imputation accuracy, minimizes estimation bias, and improves predictive performance, particularly in industrial datasets with high proportions of missing values. The findings demonstrate that the framework provides a reliable and efficient solution for managing incomplete data, thereby supporting robust industrial analytics and informed decision-making.*

Keywords— *Multiple Imputation, Missing Not at Random (MNAR), Data Augmentation, Selection Model, Semi-parametric Estimation, Darboux Variate, Missing Data Analysis, Industrial Data Analytics, Predictive Modeling, Statistical Inference.*

Date of Submission: 01-08-2026 Date of acceptance: 07-08-2026

I. INTRODUCTION

The rapid growth of data-driven research has transformed decision-making across diverse fields, including healthcare, finance, social sciences, business analytics, engineering, and artificial intelligence. The quality of decisions derived from statistical analysis and data mining depends largely on the completeness and reliability of the available data. However, real-world datasets are rarely complete, as missing values frequently arise during data collection due to participant nonresponse, equipment failure, data entry errors, attrition in longitudinal studies, privacy concerns, or incomplete administrative records. Even a small proportion of missing observations can reduce statistical efficiency, introduce bias, and compromise the validity of analytical results if not handled appropriately.

Missing data mechanisms are generally classified into three categories: Missing Completely at Random (MCAR), Missing at Random (MAR), and Not Missing at Random (NMAR). Among these, the MAR assumption has received widespread attention because it enables the application of likelihood-based estimation and conventional multiple imputation techniques to obtain unbiased parameter estimates. Nevertheless, the MAR assumption is often difficult to justify in practical applications. In many real-world situations, the probability that an observation is missing depends directly on the unobserved value itself. For example, individuals with exceptionally high incomes may intentionally avoid reporting their earnings, patients experiencing severe health conditions may discontinue medical studies, and respondents with sensitive personal information may decline to answer specific survey questions. Such situations correspond to the NMAR mechanism, where standard methods developed under the MAR assumption can produce biased estimates, distorted confidence intervals, and misleading statistical conclusions.

Addressing NMAR data remains one of the most challenging problems in modern statistics because the missing-data mechanism cannot be ignored during the estimation process. To overcome this limitation,

researchers have proposed several modelling frameworks that jointly characterize the outcome variable and the response mechanism. Selection models, pattern-mixture models, shared-parameter models, Bayesian inference, and semiparametric estimation methods have significantly advanced the analysis of nonignorable missing data. These approaches explicitly account for the dependence between missingness and unobserved outcomes, thereby improving the accuracy and reliability of statistical inference. Recent developments have also combined iterative multiple imputation with data augmentation algorithms, enabling efficient estimation while accommodating complex data structures and reducing reliance on restrictive distributional assumptions. Motivated by these challenges, the present study investigates a semiparametric selection-model approach for handling NMAR data through multiple imputation.

The proposed framework integrates data augmentation with flexible estimation procedures to recover missing observations while minimizing model dependence. Unlike conventional imputation methods that rely on strong assumptions regarding the complete-data distribution, the proposed methodology focuses on modelling the observed responses and estimating the missing-data mechanism simultaneously. This strategy enhances robustness against model misspecification and improves estimation accuracy across a wide range of practical scenarios. The major contribution of this research is the development and evaluation of a robust multiple imputation framework capable of producing reliable parameter estimates under nonignorable missing-data conditions. The proposed methodology is assessed through comprehensive simulation experiments and empirical analysis to examine its statistical accuracy, bias, efficiency, and robustness.

The findings demonstrate that incorporating the missing-data mechanism into the estimation process substantially improves inferential performance compared with traditional approaches that assume missingness is ignorable. Consequently, this study contributes to the advancement of missing-data methodology by providing an effective and flexible framework for analysing incomplete datasets encountered in contemporary scientific and data-driven research.

II. LITERATURE REVIEW

Missing data remain a persistent challenge in statistical modelling, data mining, and machine learning because incomplete observations can reduce estimation efficiency, increase uncertainty, and introduce systematic bias into analytical results. As modern datasets continue to grow in size and complexity, developing reliable methods for handling missing information has become an important area of statistical research. Numerous techniques have been proposed over the past four decades to improve the quality of inference from incomplete datasets while preserving the underlying characteristics of the observed data [1], [2].

The concept of Multiple Imputation (MI) represented a significant advancement in missing-data analysis by replacing each missing value with several plausible estimates rather than a single deterministic value. This approach allows the uncertainty associated with missing observations to be incorporated into subsequent statistical inference, thereby producing more reliable parameter estimates and confidence intervals [1], [2]. Later studies established the theoretical framework for categorizing missing-data mechanisms into Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Among these mechanisms, MNAR presents the greatest analytical difficulty because the probability of missingness depends directly on the unobserved values, making conventional likelihood-based estimation methods inappropriate without explicitly modelling the response process [2].

To address the challenges associated with nonignorable missing data, several researchers proposed statistical models that jointly characterize both the outcome variable and the missing-data mechanism. The sample selection model introduced an effective strategy for correcting selection bias by incorporating the response probability into the estimation procedure [3]. Subsequently, pattern-mixture models provided an alternative framework by modelling respondents and nonrespondents separately, allowing greater flexibility in analyzing longitudinal and multivariate incomplete datasets [4], [5]. These modelling strategies demonstrated substantial improvements over complete-case analysis and single imputation by reducing estimation bias and improving inferential accuracy.

Bayesian computation has also played a fundamental role in advancing missing-data methodology. The development of Markov Chain Monte Carlo (MCMC) methods enabled efficient estimation for complex statistical models involving latent variables and incomplete observations [6]. Adaptive Rejection Sampling further improved computational efficiency for Bayesian estimation by facilitating Gibbs sampling in models with log-concave posterior distributions [7]. These developments contributed directly to the widespread adoption of Data Augmentation, an iterative estimation framework that alternately updates missing values and model parameters until convergence is achieved, resulting in improved stability and estimation accuracy [8].

Recent studies have focused on developing computationally efficient methods specifically designed for MNAR datasets. Semiparametric estimation approaches reduce reliance on restrictive distributional assumptions while maintaining statistical consistency under nonignorable missingness [9]. Fractional imputation techniques

have also demonstrated improved computational performance without compromising estimation quality [10]. Furthermore, integrating multiple imputation with selection models has produced more robust estimation procedures capable of addressing complex healthcare, industrial, and socioeconomic datasets characterized by informative nonresponse [11], [12].

Another important research direction has emphasized reducing nonresponse bias through calibration weighting and response propensity modelling. Calibration estimators utilize auxiliary information to improve population estimates under incomplete response, while response propensity models estimate participation probabilities to compensate for systematic nonresponse [13], [14]. These approaches have proven particularly valuable in large-scale surveys and official statistics, where missing observations frequently arise because of participant refusal or operational constraints. Propensity-score adjustment techniques have further enhanced estimation accuracy under informative sampling designs by reducing bias caused by nonrandom response mechanisms [15], [16].

With the emergence of big data analytics, artificial intelligence, and Industrial Internet of Things (IIoT) applications, missing-data problems have become increasingly common in healthcare, manufacturing, finance, and smart infrastructure. Modern industrial systems continuously generate high-dimensional sensor data that often contain incomplete observations resulting from communication failures, hardware malfunctions, or data transmission errors. Robust multiple imputation methods integrated with data augmentation algorithms have therefore become essential for maintaining data integrity, improving predictive modelling, and supporting reliable decision-making in real-world environments [11], [12].

Overall, the existing literature demonstrates that integrating Multiple Imputation, Data Augmentation, and selection-model-based estimation provides a comprehensive framework for analyzing NMAR datasets. Compared with traditional deletion or single-imputation techniques, these methods effectively account for uncertainty associated with missing observations, reduce estimation bias, improve statistical efficiency, and generate more reliable statistical inference. Building upon these advances, the present study adopts a semiparametric selection-model approach combined with data augmentation to improve multiple imputation performance under nonignorable missing-data conditions.

III. NONIGNORABLE MISSING DATA AUGMENTATION METHOD

A. Parametric approach

The proposed **parametric outcome-based imputation technique** utilizes the estimated relationship between respondent and nonrespondent outcome distributions to recover missing observations. By explicitly accounting for distributional differences, the method improves the accuracy of imputed values, minimizes estimation bias, and enhances the robustness of statistical analyses involving incomplete data.

$$y_i^* \sim \int (x|y, \delta = 0) \neq \int (x|y, \delta = 1)$$

The notation (y_i^{*}) represents the imputed value assigned to the missing response of the (i^{th}) observation after the imputation procedure is completed. Although rejection sampling is theoretically applicable for generating missing observations, its practical implementation is often limited because the odds function of the nonresponse probability, (O(x_i, y; \phi)), may not possess a finite upper bound. Consequently, the sampling process becomes computationally inefficient and unsuitable for datasets exhibiting nonignorable missingness.

To overcome this limitation, the **Parametric Fractional Imputation (PFI)** approach proposed by Kim (2011) is employed as an efficient alternative. Unlike conventional rejection sampling, PFI eliminates the need for repeated iterative sampling while maintaining estimation accuracy. Initially, (L) candidate values are generated for each missing observation using the estimated outcome model constructed from the respondent data. These candidate values are subsequently assigned fractional weights based on their corresponding estimated response probabilities. The weighted candidate values collectively represent the uncertainty associated with the missing observation, enabling efficient estimation while reducing computational complexity and improving the robustness of the imputation process.

The fractional weight is expressed as:

$$w(\,ij)\hat{}* = O\ (x1i, y\ (j)\ i\ ;\ \varphi)/\sum L\ j = 1\ O(x1i, y\ (j)\ i\ ;\ \varphi)$$

Here, y (j) i is represents the jth candidate imputed value generated from the estimated respondent outcome model. The normalized fractional weights determine each candidate's contribution, ensuring statistically consistent and reliable imputation while appropriately accounting for nonignorable missing data.

B. Semi parametric approach

The semiparametric approach improves robustness by reducing sensitivity to model misspecification while preserving *flexibility in estimating missing values*. It combines kernel weighting with response probabilities, expressed as:

$$x^*_{i,semi} = \sum_{j=1}^{n} \delta_j K_h(y_j, y_i) O(y1_i, x_i^{(j)}; \emptyset) x_i^{(j)} / \sum_{j=1}^{n} \delta_j K_h(y_j, y_i) O(y1_i, x_i^{(j)}; \emptyset),$$

In this formulation, (K_h(\cdot)) denotes a symmetric kernel function with bandwidth (h), where the bandwidth parameter governs the degree of smoothing applied during the estimation process. Although direct substitution of predicted values provides a simple imputation strategy, it often fails to adequately represent the inherent variability of the missing data, leading to an underestimation of both within-imputation and between-imputation uncertainty. To preserve the natural variability within each imputed dataset, random residual components are incorporated into the imputed values. Furthermore, bootstrap resampling is employed across multiple imputed datasets to capture the variability arising from the imputation process itself. The combined use of residual-based adjustment and bootstrap sampling produces statistically valid variance estimates and improves the reliability of subsequent parameter estimation and inference.

In each iteration, the missing value is imputed as, $x_i^{**} = x^*_{i,semi} + \hat{e_i}$, where $\hat{e_i}$ Randomly drawn from the estimated residual values $\hat{e_j}$, where $\hat{e_j} = x_j - \hat{x}_{i,semi,j} = 1, \ldots, n_R$

Overall multiple imputation estimate is calculated as:

$$\hat{\mu}y = \frac{1}{M} \sum_{k=1}^{M} \hat{\mu}y^{(k)}$$

The variance of the mean estimator is obtained using Rubin's (1987) variance formulation that is,

$$\hat{V}y = W_M + (1 + M^{-1}) B_M ,$$

Variance estimation plays a critical role in evaluating the precision and reliability of parameter estimates obtained from incomplete data. Although linearization-based approaches have been widely adopted for estimating sampling variance under complex survey designs, their implementation often becomes computationally demanding when the estimation procedure involves nonlinear models, multiple imputation, or nonignorable missing-data mechanisms. In particular, the derivation of analytical variance expressions requires repeated approximation of complex estimating equations, making these methods less practical for high-dimensional datasets and iterative estimation algorithms. Consequently, direct variance estimation techniques have gained increasing attention because they provide a computationally efficient framework while maintaining satisfactory statistical accuracy. By avoiding extensive analytical derivations, these methods simplify implementation and are readily adaptable to modern imputation procedures, including semiparametric and Bayesian frameworks.

Consider the set $S = \{P \in [M, N] \mid [M, P] \text{ contains only finitely many elements of } (Rn)\}$.The point **M** belongs to **S** because the interval **[M, M]** contains at most one element of the sequence. Furthermore, if a point **P** belongs to **S**, then every point between M and **P** also belongs to **S**, since each corresponding subinterval contains only finitely many elements of the sequence. As a result, **S** forms a continuous interval starting from M and extending towards a limiting point. Let R=sup(S), where R represents the least upper bound of **S**. For any **ε > 0**, the interval [M, R − ε] contains only finitely many terms of the sequence, whereas the interval [M, R + ε] contains infinitely many terms. Consequently, every neighbourhood (R−ε,R+ε) contains infinitely many elements of the **sequence** (R_n)**.** Therefore, R **is** an accumulation point of the sequence, which guarantees the existence of a convergent subsequence. This property provides an important mathematical foundation for the convergence and stability of statistical estimation procedures.

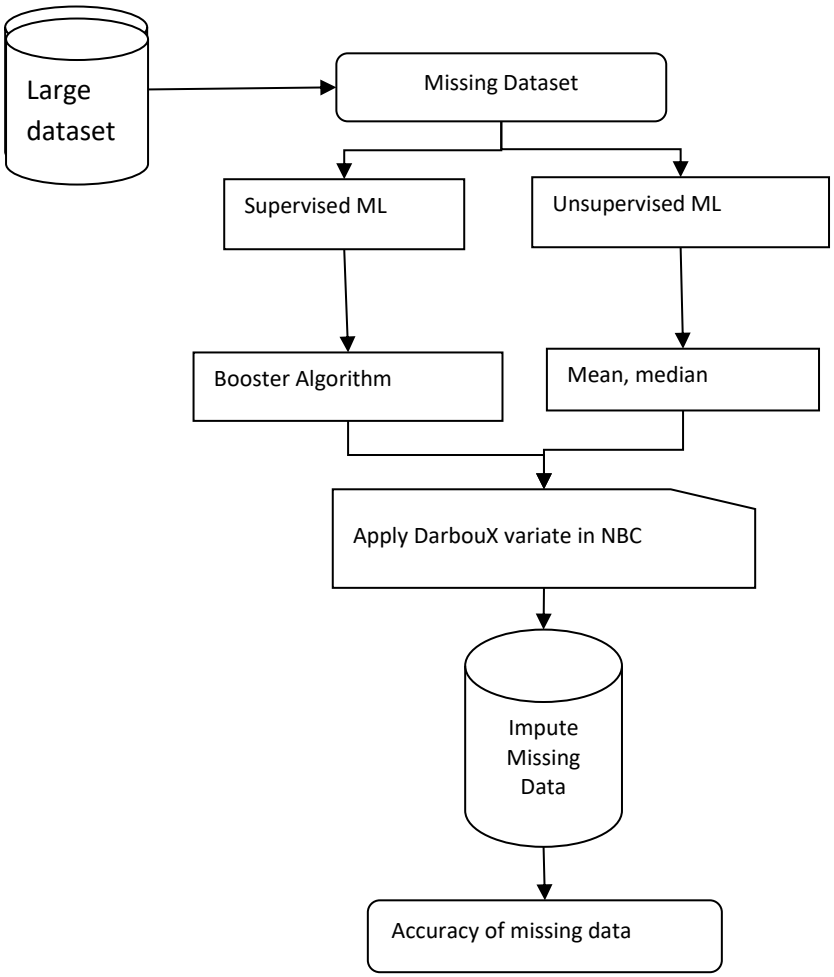

Fig.1. Enhanced missing data accuracy using Darboux variates and Naïve Bayes Classifier

The existence of a convergent subsequence for every bounded sequence provides an important theoretical foundation for statistical estimation and data analysis. This property extends naturally to higher-dimensional spaces, including, R^n where closed and bounded sets also satisfy compactness conditions. In missing data analysis, these mathematical principles support the development of stable imputation techniques. One commonly used approach is the Naïve Bayes Classifier (NBC)**,** where the attribute containing missing values is treated as the target variable, while the remaining attributes serve as predictor variables. The classifier is trained using complete observations and then estimates missing values based on conditional probability relationships among the variables. To improve the reliability of the imputed values, lower and upper bounds of the observed data are considered, restricting predictions to a feasible range. This bounded estimation reduces unrealistic imputations and enhances numerical stability. Consequently, combining NBC with boundedness and convergence principles produces consistent, accurate, and theoretically sound estimates for datasets containing missing values.

EXPERIMENTAL RESULTS

The experimental evaluation was conducted using the University dataset obtained from the UCI Machine Learning **Repository**. The dataset consists **of** 265 instances described **by** 13 attributes, including both categorical and numerical variables. To assess the effectiveness of the proposed imputation approach, a complete version of the dataset containing no missing values was initially selected. Artificial missing values were then introduced randomly at different proportions, ranging from 5% to 25%**,** to simulate various levels of data incompleteness. The missing values were distributed across multiple attributes to generate diverse experimental conditions. The proposed classification-based imputation method was subsequently applied to estimate the missing entries. Its performance was evaluated by comparing the imputed values with the original dataset, enabling an assessment of the method's accuracy, robustness, and consistency under different missing data rates.

Data Set Characteristics:	**Multivariate**	**Number of Instances:**	265
Attribute Characteristics:	Categorical, Integer	Number of Attributes**:**	13
Associated Tasks:	Classification	Missing Values	Yes

TABLE 1: Dataset Used for Analysis

The following figure compares the performance of different missing value imputation methods. It evaluates supervised techniques, including Naïve Bayes, Booster Algorithm, and the proposed NBC–Darboux Variate approach. These methods are compared with Mean, Median, and Standard Deviation-based imputation techniques. The comparison measures their ability to estimate missing values accurately. The results demonstrate the effectiveness of each method

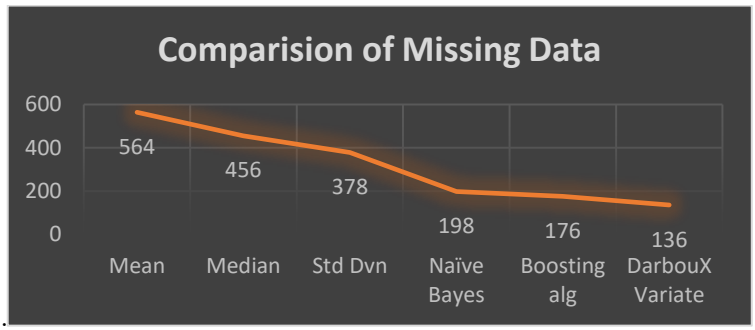

Fig.2. Missing Values imputation in Original Dataset

The below figure presents the percentage of missing values introduced at **5%, 10%, 15%, 20%, and 25%**. It compares supervised approaches, including **NBC, Boosting Algorithm, and NBC–Darboux Variate**, with unsupervised methods such as **Mean, Median, and Standard Deviation**. The evaluation is performed across multiple dataset attributes. The comparison demonstrates the effectiveness of each imputation technique under different levels of missing data

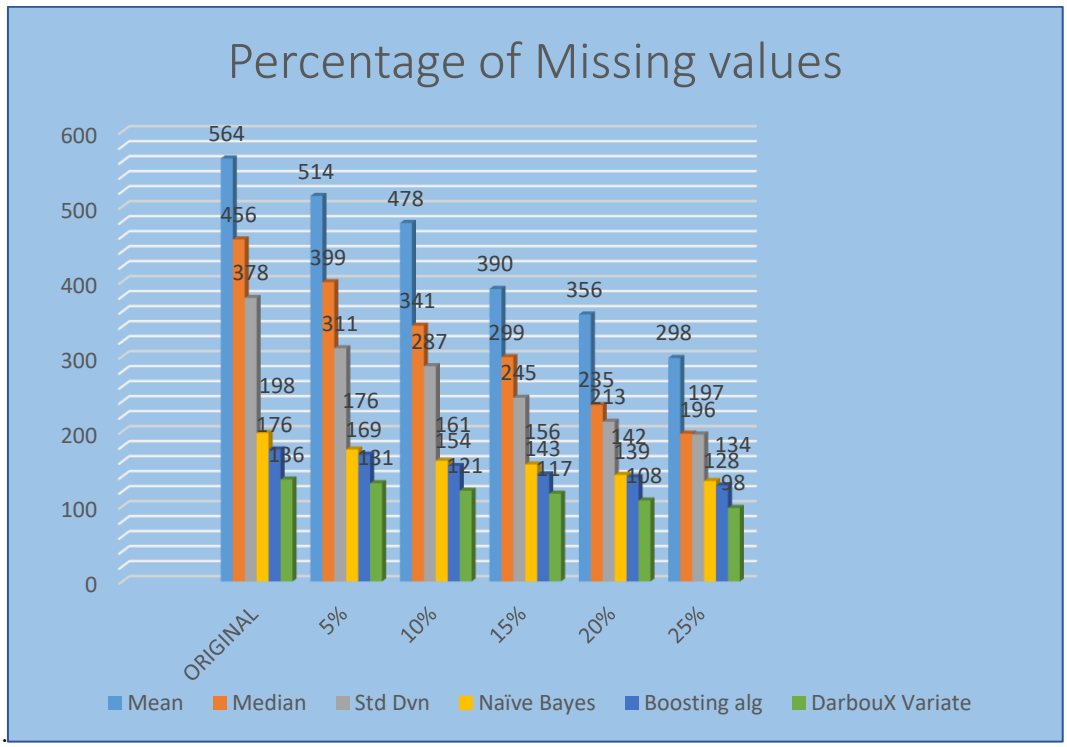

Fig 3. Percentage Distribution of Missing Values

IV. CONCLUSION

Nonignorable missing data continue to present a significant challenge in statistical inference because the probability of missingness depends directly on unobserved values. Under such circumstances, conventional imputation techniques developed for ignorable missing-data mechanisms often produce biased parameter estimates and unreliable statistical conclusions. Although existing frameworks, including selection models and pattern-mixture models, have been widely applied to address nonignorable nonresponse, their performance is frequently influenced by stringent assumptions regarding the response mechanism and the underlying outcome distribution. Any violation of these assumptions may adversely affect estimation accuracy and reduce the robustness of subsequent analyses. To address these limitations, this study introduces a semiparametric multiple imputation framework that jointly models the respondents' outcome distribution and the response mechanism through the incorporation of a nonresponse instrumental variable. The proposed methodology employs a data augmentation strategy that integrates fractional imputation with multiple imputation to generate statistically consistent estimates for incomplete observations. Instead of relying on deterministic replacement, multiple candidate values are first generated from the estimated outcome model, after which probability-based fractional weights are assigned according to the estimated response propensity. Missing observations are then imputed through weighted probability sampling, allowing the uncertainty associated with nonresponse to be naturally incorporated into the estimation process. This framework substantially reduces computational complexity while maintaining estimation efficiency and improving the statistical validity of the imputed datasets.

A key advantage of the proposed approach is its adaptability to diverse data structures and modelling assumptions. Because the outcome model is estimated exclusively from observed responses, its adequacy can be verified using standard goodness-of-fit procedures or extended through semiparametric and nonparametric estimation techniques when strict distributional assumptions are difficult to justify. Consequently, the proposed framework demonstrates increased robustness against moderate model misspecification while producing more reliable parameter estimates and confidence interval coverage under nonignorable missing-data conditions.

In addition to model-based imputation methods, classification-based techniques have also been investigated for recovering incomplete observations. One such approach is the Darboux Variate Naïve Bayes Imputation Classifier, which estimates missing values by identifying influential predictor variables and determining an appropriate sequence for the imputation process before applying the Naïve Bayes classification algorithm. The incorporation of Darboux variates utilizes the infimum and supremum characteristics of ordered data sequences to improve the stability of probability estimation and reduce uncertainty during classification. This strategy enhances the robustness of the imputation procedure, particularly when incomplete observations exhibit nonlinear or irregular patterns.

Despite its computational simplicity and scalability, the conventional Naïve Bayes classifier assumes conditional independence among predictor variables, an assumption that is often violated in practical datasets. Furthermore, the occurrence of previously unseen attribute combinations may lead to zero posterior probabilities, thereby reducing classification performance. The integration of Darboux variates alleviates these shortcomings by incorporating interval-based information into the probability estimation process, improving predictive accuracy and enhancing the reliability of missing-value recovery. Overall, the combination of semiparametric multiple imputation, fractional weighting, data augmentation, and Darboux variate-assisted probabilistic classification provides a comprehensive and flexible framework for analysing complex incomplete datasets, yielding improved estimation accuracy, reduced bias, and more dependable statistical inference across a broad range of real-world applications.

REFERENCES

[1]. G. E. P. Box and G. G. Tiao, *Bayesian Inference in Statistical Analysis*. New York, NY, USA: Wiley, 1992. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118033197

[2]. J. R. Carpenter, M. G. Kenward, and I. R. White, "Sensitivity analysis after multiple imputation under missing at random: A weighting approach," *Statistical Methods in Medical Research*, vol. 16, pp. 259–275, 2007. Available: https://doi.org/10.1177/0962280206075303

[3]. T. Chang and P. S. Kott, "Using calibration weighting to adjust for nonresponse under a plausible model," *Biometrika*, vol. 95, no. 3, pp. 555–571, 2008. Available: https://doi.org/10.1093/biomet/asn028

[4]. G. B. Durrant and C. Skinner, "Using data augmentation to correct for non-ignorable non-response when surrogate data are available: An application to the distribution of hourly pay," *Journal of the Royal Statistical Society: Series A*, vol. 169, no. 3, pp. 605–623, 2006. Available: https://doi.org/10.1111/j.1467-985X.2006.00415.x

[5]. J.-E. Galimard, S. Chevret, C. Protopopescu, and M. Resche-Rigon, "A multiple imputation approach for MNAR mechanisms compatible with Heckman's model," *Statistics in Medicine*, vol. 35, pp. 2907–2920, 2016. Available: https://doi.org/10.1002/sim.6903

[6]. A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990. Available: https://doi.org/10.1080/01621459.1990.10476213

[7]. W. R. Gilks and P. Wild, "Adaptive rejection sampling for Gibbs sampling," *Applied Statistics*, vol. 41, no. 2, pp. 337–348, 1992. Available: https://doi.org/10.2307/2347565

[8]. C. Giusti and R. J. A. Little, "An analysis of nonignorable nonresponse to income in a survey with a rotating panel design," *Journal of Official Statistics*, vol. 27, no. 2, pp. 211–229, 2011.
Available: https://www.scb.se/en/services/statistical-programmes-for-research/journal-of-official-statistics/
[9]. J. S. Greenlees, W. S. Reece, and K. D. Zieschang, "Imputation of missing values when the probability of response depends on the variable being imputed," *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 251–261, 1982. Available: https://doi.org/10.1080/01621459.1982.10477767
[10]. J. J. Heckman, "Sample selection bias as a specification error," *Econometrica*, vol. 47, no. 1, pp. 153–161, 1979. Available: https://doi.org/10.2307/1912352
[11]. J. Im and S. Kim, "Multiple imputation for nonignorable missing data," *Journal of the Korean Statistical Society*, vol. 46, no. 4, pp. 583–592, 2017.
Available: https://doi.org/10.1016/j.jkss.2017.05.002
[12]. J. K. Kim and C. L. Yu, "A semi-parametric estimation of mean functionals with non-ignorable missing data," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 157–165, 2011.
Available: https://doi.org/10.1198/jasa.2011.tm09466
[13]. J. K. Kim, "Parametric fractional imputation for missing data analysis," *Biometrika*, vol. 98, no. 1, pp. 119–132, 2011. Available: https://doi.org/10.1093/biomet/asq073
[14]. P. S. Kott and T. Chang, "Using calibration weighting to adjust for nonresponse and coverage errors," *Journal of the American Statistical Association*, vol. 105, no. 491, pp. 1265–1275, 2010.
Available: https://doi.org/10.1198/jasa.2010.tm09599
[15]. R. J. A. Little, "Modeling the drop-out mechanism in longitudinal studies," *Journal of the American Statistical Association*, vol. 90, no. 431, pp. 1112–1121, 1995.
Available: https://doi.org/10.1080/01621459.1995.10476615
[16]. R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. New York, NY, USA: Wiley, 2002. Available: https://doi.org/10.1002/9781119013563
[17]. R. J. A. Little and Y. X. Wang, "Pattern-mixture models for multivariate incomplete data with covariates," *Biometrics*, vol. 52, no. 1, pp. 98–111, 1996.
Available: https://doi.org/10.2307/2533169
[18]. M. Peress, "Correcting for survey nonresponse using variable response propensity," *Journal of the American Statistical Association*, vol. 105, no. 492, pp. 1418–1430, 2010.
Available: https://doi.org/10.1198/jasa.2010.ap09415
[19]. J. Qin, D. Leung, and J. Shao, "Estimation with survey data under nonignorable nonresponse or informative sampling," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 193–200, 2002.
Available: https://doi.org/10.1198/016214502753479275
[20]. M. K. Riddles, J. K. Kim, and J. Im, "Propensity-score-adjustment method for nonignorable nonresponse," *Journal of Survey Statistics and Methodology*, 2016.
Available: https://doi.org/10.1093/jssam/smv040
[21]. D. B. Rubin, "Formalizing subjective notions about the effect of nonrespondents in sample surveys," *Journal of the American Statistical Association*, vol. 72, no. 359, pp. 538–543, 1977.
Available: https://doi.org/10.1080/01621459.1977.10480610
[22]. D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*. New York, NY, USA: Wiley, 1987. Available: https://doi.org/10.1002/9780470316696