# Machine Learning Approaches for Prediction of Fertility Determinants in Bangladesh: evidence from the BDHS 2017-18 data

**Dr. Md Jamal Uddin**
Shahjalal University of Science and Technology
**Dr. Ahmad Kabir**
Shahjalal University of Science and Technology
**Shayla Naznin** ( shayla.naznin@gmail.com )
Mawlana Bhashani Science and Technology University

**Research Article**

**Keywords:** Fertility, Boruta algorithm, Machine Learning, SVM, Best Model

**Posted Date:** February 15th, 2024

**DOI:** https://doi.org/10.21203/rs.3.rs-3934391/v1



**Additional Declarations:** No competing interests reported.

## Abstract

# Background

Fertility is a social indicator that represents the country's growth and economic sustainability. The fertility rate of a country refers to number of average children born to a woman during her lifetime. It is an important demographic indicator that influences population dynamics, economic growth, social welfare, and public policy. This research leverages advanced machine learning methodologies to achieve more precise predictions of fertility and fertility determinants in Bangladesh.

# Methods

The dataset utilized in this study was sourced from the Bangladesh Demographic Health Survey (BDHS) conducted in the year 2017–18. Python 3.0 programming language were used to implement and test the machine learning (ML) models such as Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, LightGBM and Neural Network (NN). We have used Boruta algorithm of Feature selection with R programming language packages. Conventional methods were analyzed using SPSS Version 25 and R programming language. The predictive models performance was evaluated and compared with the metrics such as macro average and weighted average of the Confusion Matrix, Accuracy, F1 Score, Precision, Recall, Area Under the Receiver Operating Characteristics Curve (AUROC) and K-fold cross-validation.

# Results

We preferred with the Support Vector Machine (SVM) model of fertility in Bangladesh with macro average recall (93%), precision (89%), F1 score (90%) in addition with weighted average recall (97%), precision (96%), F1 score (96%) K-fold accuracy (95.9%). Our predictive models showed that Access to mass media, Husband/partner's education level, Highest educational level, Number of household members, Body Mass Index of mother, Number of living children and Son or daughter died stand out as the key determinants influencing fertility in Bangladesh.

# Conclusions

In the realm of constructing advanced predictive models, Machine Learning methods surpass conventional statistical approaches in classifying concealed information. In our Study the Support Vector Machine (SVM) emerged as the top-performing model for fertility prediction in Bangladesh.

## Introduction

Fertility is a vital indicator of country's standard of living or development of a population. Fertility is changing day by day since 1975. Numerous studies have indicated that the widespread availability of family planning services along with factors such as women's educational achievements and urbanization plays a crucial role in influencing women's fertility. Bangladesh has experienced decreasing the Total Fertility Rate from 6.3 births per woman in 1975 to 2.3 births per woman in 2017-18 period. But, the Total Fertility Rate (TFR) remained elevated in comparison to other developing countries.

A study by Roy [1] find that women's fertility is significantly impacted by education ($p < 0.001$). The likelihood of having no children was six times higher among individuals with secondary education compared to those who never attended school. Urban women exhibited a more pronounced manifestation of this trend compared to their rural counterparts. Ahbab [2] find that death of last child, gender preference, abortions, couples educational level, mother's participation in labor force plays a significant role in fertility. In addition to Chen's study [3] reveals that economic development indicators significantly contribute to the decline in fertility in Bangladesh, with certain climate change parameters and crop production emerging as noteworthy factors. A research by Islam [4] examined the role of proximate determinants was examined, revealing that contraception has become the most significant factor in reducing fertility rates in Bangladesh in recent years. Notably, its impact is most pronounced among middle-aged and older women. Also another research paper from Islam [5] conducted an analysis and the results indicate fertility declined significantly with women's education in addition to the other factors for instance region, place of residence and household wealth status plays important roles also. All factors examined in the study displayed a significant influence on the number of children ever born. A research from Adhikari [6] conducted and find that age at first marriage, literacy status, perceived ideal number of children, mass media exposure, wealth status, and child-death experience by mothers are paying a vital role on fertility differentials in Nepal. A study was conducted Bora [7] it was suggested that the more women's education and the resulting adoption of family sizes is decreasing could be the key driving force behind the remarkable decline in fertility rates in Bangladesh. A research from Haq [8] showed that an increase in age at marriage significantly reduced the total fertility of women in Bangladesh. This analyses also showed negative association between mean fertility and educational level of the respondents. Research by Hahn [9] suggests that women with higher education experiencing less fertility rates, exhibit greater utilization of maternal health care services, and witness improved health outcomes in their children compared to women with lower levels of education. A study was conducted by Kamal [10] focus was on investigating the influence of education on fertility utilizing from the BDHS 2007 data in Bangladesh. These findings is showing negative association between women's education and cumulative fertility. A concerning observation is the notable percentage of Bangladeshi women exhibiting high-risk fertility behavior, signaling a cause for alarm. According to findings from Howlader [11], the research is showing 67.7% women exhibited high risk fertility behavior. Among them, 45.6% faced a single risk, while 22.1% encountered multiple higher risks. Furthermore, the study suggests an association between higher education levels of women and their partners with a reduction in high-risk fertility behavior. Zelalem [12] shown on a research study examining the levels and fertility patterns among women within the Kersa District of East Ethiopia. Higher fertility rates were observed among rural illiterate women compared to urban literate women, as revealed by the author's analysis using follow-up data from 2008 to 2012. According to Abedin [13] the research highlights the significant impact of women's decision-making is influencing on fertility outcomes.

Some of the studies highlight the applications of machine learning methods in various field of demography [16–20]. In a study shows that a woman's age, education level, occupation, and location are significant determining factors for the survival of a child during birth delivery [23]. In another study shows that various factors such as age, ethnicity, gender, household registration, occupation, education, duration of residence, housing, migration scope, economic status and health services access are influenced in fertility behavior [24]. In a study employing Neural Network method to find the determinant of fertility and his findings was religion, status of education, wealth index, current age of the respondent and contraceptive methods are the key determinants of fertility [25].

Although many of works were done previous studies for finding determinants and prediction in many field but there very few number of research were done to find the fertility determinant and prediction using Machine Learning approach. This study endeavors to (i) Utilize Machine Learning methodologies to discern the factors influencing fertility in Bangladesh and (ii) Apply Machine Learning approaches to predict fertility in Bangladesh using BDHS 2017–18 data.

# Methods and Materials

### Data

This study utilizes Bangladesh Demographic and Health Survey (BDHS) conducted in 2017-18 data which is part of an ongoing series initiated in 1993. The nationally representative surveys employed covering all administrative divisions in Bangladesh. The BDHS ensured consistency in comparing different demographics data by employing nearly identical questionnaires over time. Comprehensive details of sampling and survey methodology regarding the BDHS have been previously documented [14] and all survey data considered secondary data are publicly accessible [15]. For this research we specifically extracted information from the most recent BDHS (2017-2018) datasets with a focus on female respondents excluding temporary (de jure) residents and cases with missing values. The dataset comprising 20,127 weighted observations serves as a robust foundation for our analysis.

### Ethical approval

Bangladesh Demographic and Health Survey (BDHS) secondary data were used in this study, this is a nationally representative survey conducted in Bangladesh at regular intervals. The BDHS collects crucial information on health, family planning, and socio-demographic factors. It is carried out by the National Institute of Population Research and Training (NIPORT) under the Ministry of Health and Family Welfare, with technical support from ICF International. Permission for the survey protocol was secured from the ORC Macro (Macro International Inc.) Institutional Review Board and the National Research Ethics Committee in Bangladesh. The study relied on existing public domain survey datasets freely available online. Before analysis, all identifying information of the respondents was removed. The dataset for this research was acquired after obtaining permission by the authors from the Demographic and Health Surveys (DHS) program.

### Target variable

The primary target variable in the study was children ever born (CEB), representing information of the total number of children born up to the survey date.

### Independent variables

The independent variables are shown in the Table-1

Table-1: Independent variables

| Variables | Variables Coding |
| --- | --- |
| Division | 0= Barishal, 1= Chattogram, 2= Dhaka, 3= Khulna, 4= Rajshahi, 5= Mymensingh, 6= Rangpur, 7= Sylhet |
| Type of place of residence | 1= Rural, 0= Urban |
| Religion | 1=Muslim, 2=Non-Muslim |
| Wealth index | 0=Poorest, 1=Poorer, 2=Middle, 3=Richer, 4=Richest |
| Mother's educational level | 0=No, 1=education, 2=Primary, 3=Secondary, 4=Higher |
| Father's education level | 0=No, 1=education, 2=Primary, 3=Secondary, 4=Higher |
| Father's occupation | 1=Agriculture, 2=Worker or Labor, 3=Professional Worker, 4=Business & Others |
| Mother's occupation | 0= Housewife, 1=Worker or Labor, 2=Professional Worker, 3=Business & Others |
| Type of Toilet Facilities | 1=Toilet With Flush, 2= VIP latrine, 3=Pit Latrine, 4=Hanging toilet and Other |
| Sources of Drinking Water | 1=Other than Tubewell or Borehole, 2=Tubewell or Borehole |
| Number of household members | 1=1-3 Member, 2=4-6 Member, 3=Over 6 Member |
| Type of Cooking Fuel | 1= Kerosene or Natural Gas, 2=Wood, 3=Agricultural Crop, 4=Animal Dung and Others |
| Number of living children | 0=No Child, 1=One Child, 2=Two Child, 3=Three Child, 4=Four Child, 5=Five or more Children |
| Body Mass Index (BMI) | 1=Thin(<18.5), 2=Normal (18.5-24.99), 3=Overweight (25-29.99), 4=Obese(>=30) |
| Age at first Marriage | 1=Below 18years old, 2=18 years and above |
| Contraceptive use and intention | 1=Using modern method, 2=Using traditional method, 3=Non-user - intends to use later, 4=Does not intend to use |
| Son or daughter died | 0=No, 1=Yes |
| Access to Media | 0=No, 1=Yes |
| Sex of household head | 1=Male, 2=Female |
| Ideal number of | 1=0-2 Children, 2=3-4 Children, 3=Over 4 Children |

| | |
|---|---|
| children | |
| Currently breastfeeding | 0=No, 1=Yes |
| Desire for more children | 1=Wants within 2 years, 2=Wants after 2+ years, 3=Wants, unsure timing/Undecided, 4=Wants no more, 5=Sterilized (respondent or partner)/Declared infecund |
| Father's desire for children | 1=Both want same, 2=Husband wants more, 3=Husband wants fewer, 4=Don't know |

### Statistical analysis

This study aims to explore risk factors associated with fertility by employing various machine learning classification models namely Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, LightGBM and Neural Network (NN). The dataset was divided into an 80% training set and a 20% test set. The training set was used to train the machine learning models, while the test set evaluated their performance. Subsequently, the entire dataset was used to predict with the trained models. The evaluation of predictive models was conducted and compared with metrics such as the macro average and weighted average of the Confusion Matrix, Accuracy, F1 Score, Precision, Recall, Area Under the Receiver Operating Characteristics Curve (AUROC) and K-fold cross-validation.

The chi-square tests were used to assess the variables which is significant to fertility, facilitating a comparison between machine learning and traditional approaches. Furthermore, the Boruta algorithm was utilized to pinpoint the crucial features associated with fertility. The study utilized Python version 3.0 for machine learning methods, in the R programming language, the Boruta package were employed to select the best features. Furthermore, SPSS Version 25 was used for calculating bivariate analysis.

## Results

### Descriptive Statistics

Table-2 illustrates the frequency distribution of the number of children ever born to mothers, along with the corresponding chi-square values and p-values. According to the table, the Dhaka division has the highest proportion of mothers (25.45%), while the Barishal division has the lowest percentage, with only 5.59% of mothers. Khulna notably stands out with 10.15% having no children 77.14% with 1 to 3 children and 12.71% with 4 or more children. Dhaka division 25.45% of the population with 11.73% having no children 69.82% having 1 to 3 children and 18.45% having 4 or more children. Urban areas (28.46%) have 10.96% with no children 73.24% with 1 to 3 children and 15.8% with 4 or more children. Contrastingly rural regions (71.54%) exhibit 9.7%, 67.52% and 22.77% respectively. The Muslim population founds 90.67% while non-Muslims are 9.33% on all surveyed mothers. The wealth index unveils disparities with percentages varying among wealth categories: Poorest (18.6%), Poorer (19.66%), Middle (20.17%), Richer (20.79%) and Richest (20.79%). This shows that the richest with 12.31% no children 76.26% with 1 to 3 children and 11.43% with 4 or more children. Maternal and paternal education levels show significant effect on fertility (child ever born). For instance, maternal education highlights the distribution: No education (16.56%), Primary (31.25%), Secondary (39.61%) and Higher (12.58%). Mother whose have no education tendency of taking children is more than other groups 46.74% with 4 or more children whereas 3.36% with no children 49.89% with 1 to 3 children. Father's education echoes a similar pattern: No education (21.54%), Primary (32.12%), Secondary (29.98%) and Higher (16.36%). Fathers who have no education tendency of taking children is more than the other groups 3.7% with no children 58.40% with 1 to 3 children and 37.9% with 4 or more children. Father's occupation plays a vital role on fertility agriculture (25.87%), worker/labor (47.9%), professional worker (5.52%) and business & others (20.71%). Fathers who are professional worker with 1 to 3 children (75.24%), with no children (17.88%) and other professions are showing similar results. Mother's occupation plays a vital role on fertility housewife (49.83%), worker/labor (34.34%), professional worker (12.33%) and business & others (3.5%). Mothers whose professions are housewife has 13.57% with no children 69.9% with 1 to 3 children and 16.53% with 4 or more children and business and other profession has 16.31% with no children 71.06% with 1 to 3 children and 12.62% with 4 or more children. Our study shows household flush toilet (30.95%), VIP latrine (14.86%), latrine with a pit (52.3%) and hanging toilet and other (1.9%). Toilet facilities shows impact on fertility toilet with flush 15.37% has 4 or more children whereas others option and hanging toilet 28.21% has 4 or more children. Number of household member has significant effect on fertility our result shows 1-3 member has 4 or more children (11.46%) on the other hand household has 4-6 members has no children (7.36%). Types of cooking fuel is also influence on fertility our results reveals that using natural gas/kerosene number of high fertility is less has 11.01% with no children 75.30% with 1 to 3 children and 13.69% with 4 or more children. Age of mother has significant effect on fertility here 87.73% of mothers of our study has comes from 25-29 with 1 to 3 children. In this study thin (11.55%), normal (56.47%), overweight (25.45%) and obese (6.52%) all the class shows almost similar pattern in data. The data shows a significant percentage (75.35%) marrying before 18 years old, correlating with a higher number of children born while those marrying after 18 years old (24.65%) tend to have fewer children. Respondents who experienced the death of a son or daughter exhibit a different distribution in the number of children born. Those families who have experienced son or daughter shows more (63.15%) has 4 or more children. Access to media reflects on fertility, households with access (65.98%) and without access (34.02%) number of children born. This shows that the families who have access to media they have less children. Married individuals comprise the majority (94.33%) while smaller percentages include widowed (3.05%), divorced (1.53%) and those no longer living together/separated (1.1%). These statuses exhibit diverse correlations with the number of children born. Categories indicating the ideal number of children (0-2, 3-4, Over 4) and the desire for more children (Wants within 2 years, Wants after 2+ years, Wants no more, Sterilized/Declared infecund) display varying distributions in the actual number of children born reflecting preferences and family planning practices. Categories representing marital status (Married, Widowed, Divorced, No longer living together/separated) and husband's desires for children showcase varying distributions in the number of children born, emphasizing the role of marital dynamics and spousal preferences in family size. Households led by males versus females exhibit differences in the distribution of children born, indicating potential gender-related influences on family size within different household structures.

The result showing in the table-2 that fertility determinants are significant at 5% level of confidence, Division ($\chi2=323.96$, $p<0.001$), Type of place of residence ($\chi2=121.81$, $p<0.001$), Religion ($\chi2=94.98$, $p<0.001$), Wealth index ($\chi2=464.30$, p=<0.001), Maternal educational level ($\chi2=3218.38$, $p<0.001$), Father's education level ($\chi2=1599.57$, $p<0.001$), Father's occupation ($\chi2=621.71$, $p<0.001$), Mother's occupation ($\chi2=730.96$, $p<0.001$), Type Of Toilet Facilities $\chi2=302.31$, $p<0.001$), Sources of Drinking Water ($\chi2=32.94$, $p<0.001$), Number of household members ($\chi2=517.49$, $p<0.001$), Type of Cooking fuel ($\chi2=223.78$ , $p<0.001$), Number of living children ($p<0.001$), Mother's Age ($\chi2=8545.36$, $p<0.001$), Body Mass Index ($\chi2=158.29$, $p<0.001$), Age at first Marriage ($\chi2=638.86$, $p<0.001$), Contraceptive use and intention ($\chi2=2685.50$, $p<0.001$), Son or daughter died

($\chi2=4071.07$, $p<0.001$), Access to Media ($\chi2=658.04$, $p<0.001$), Current marital status ($\chi2=306.46$, $p<0.001$), Sex of household head ($\chi2=7.78$, $p=0.020$), Ideal number of children ($\chi2=1823.94$, $p<0.001$), Desire for more children ($\chi2=6110.81$, $p<0.001$), Currently breastfeeding ($p<0.001$), Father's desire for children ($\chi2=565.93$, $p<0.001$).

**Table-2 Descriptive analysis of Background characteristics**

| Characteristics | Children Ever Born | | | | | |
|---|---|---|---|---|---|---|
| | Total | No children | 1 to 3 children | 4 or More Children | χ2 | p-value |
| | n(%) | n(%) | n(%) | n(%) | | |
| **Division** | | | | | | |
| Barishal | 1125(5.59) | 102(9.07) | 757(67.29) | 266(23.64) | | |
| Chattogram | 3622(18) | 355(9.8) | 2325(64.19) | 942(26.01) | | |
| Dhaka | 5123(25.45) | 601(11.73) | 3577(69.82) | 945(18.45) | | |
| Khulna | 2336(11.61) | 237(10.15) | 1802(77.14) | 297(12.71) | 323.96 | <0.001 |
| Mymensingh | 1546(7.68) | 160(10.35) | 1006(65.07) | 380(24.58) | | |
| Rajshahi | 2801(13.92) | 261(9.32) | 2064(73.69) | 476(16.99) | | |
| Rangpur | 2380(11.83) | 196(8.24) | 1681(70.63) | 503(21.13) | | |
| Sylhet | 1193(5.93) | 114(9.56) | 705(59.09) | 374(31.35) | | |
| **Type of place of residence** | | | | | | |
| Urban | 5729(28.46) | 628(10.96) | 4196(73.24) | 905(15.8) | | |
| Rural | 14398(71.54) | 1397(9.7) | 9722(67.52) | 3279(22.77) | 121.81 | <0.001 |
| **Religion** | | | | | | |
| Muslim | 18251(90.67) | 1889(10.35) | 12435(68.13) | 3927(21.52) | 94.98 | <0.001 |
| Non-Muslim | 1877(9.33) | 137(7.3) | 1483(79.01) | 257(13.69) | | |
| **Wealth index** | | | | | | |
| Poorest | 3743(18.6) | 277(7.4) | 2407(64.31) | 1059(28.29) | | |
| Poorer | 3956(19.66) | 349(8.82) | 2575(65.09) | 1032(26.09) | | |
| Middle | 4059(20.17) | 387(9.53) | 2828(69.67) | 844(20.79) | 464.30 | <0.001 |
| Richer | 4183(20.79) | 497(11.88) | 2917(69.73) | 769(18.38) | | |
| Richest | 4183(20.79) | 515(12.31) | 3190(76.26) | 478(11.43) | | |
| **Maternal educational level** | | | | | | |
| No education | 3333(16.56) | 112(3.36) | 1663(49.89) | 1558(46.74) | | |
| Primary | 6290(31.25) | 342(5.44) | 4148(65.95) | 1800(28.62) | 3218.38 | <0.001 |
| Secondary | 7973(39.61) | 979(12.28) | 6225(78.08) | 769(9.65) | | |
| Higher | 2531(12.58) | 592(23.39) | 1882(74.36) | 57(2.25) | | |
| **Father's education level** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| No education | 4077(21.54) | 151(3.7) | 2381(58.4) | 1545(37.9) | | |
| Primary | 6081(32.12) | 508(8.35) | 4127(67.87) | 1446(23.78) | 1599.57 | <0.001 |
| Secondary | 5675(29.98) | 694(12.23) | 4249(74.87) | 732(12.9) | | |
| Higher | 3098(16.36) | 520(16.79) | 2402(77.53) | 176(5.68) | | |
| **Father's occupation** | | | | | | |
| Agriculture | 4902(25.87) | 295(6.02) | 3103(63.3) | 1504(30.68) | | |
| Worker/Labor | 9075(47.9) | 1078(11.88) | 6441(70.98) | 1556(17.15) | 621.71 | <0.001 |
| Professional Worker | 1046(5.52) | 187(17.88) | 787(75.24) | 72(6.88) | | |
| Business & Others | 3924(20.71) | 319(8.13) | 2838(72.32) | 767(19.55) | | |
| **Mothers's occupation** | | | | | | |
| Housewife | 10026(49.83) | 1361(13.57) | 7008(69.9) | 1657(16.53) | | |
| Worker/Labor | 6909(34.34) | 283(4.1) | 4640(67.16) | 1986(28.75) | 730.96 | <0.001 |
| Professional Worker | 2480(12.33) | 265(10.69) | 1764(71.13) | 451(18.19) | | |
| Business & Others | 705(3.5) | 115(16.31) | 501(71.06) | 89(12.62) | | |
| **Type of Toilet Facilities** | | | | | | |
| Toilet With Flush | 5724(30.95) | 650(11.36) | 4194(73.27) | 880(15.37) | | |
| VIP latrine | 2748(14.86) | 252(9.17) | 1954(71.11) | 542(19.72) | 302.31 | <0.001 |
| Pit Latrine | 9673(52.3) | 745(7.7) | 6377(65.93) | 2551(26.37) | | |
| Hanging toilet and Other | 351(1.9) | 16(4.56) | 236(67.24) | 99(28.21) | | |
| **Sources of Drinking Water** | | | | | | |
| Other than Tubewell or Borehole | 195(10.36) | 1368(72.69) | 319(16.95) | 1882(100) | 32.94 | <0.001 |
| Tubewell or Borehole | 1468(8.84) | 11393(68.57) | 3753(22.59) | 16614(100) | | |
| **Number of household members** | | | | | | |
| 1-3 Member | 4258(21.16) | 601(14.11) | 3169(74.42) | 488(11.46) | | |
| 4-6 Member | 11287(56.08) | 831(7.36) | 7976(70.67) | 2480(21.97) | 517.49 | <0.001 |
| Over 6 Member | 4581(22.76) | 593(12.94) | 2773(60.53) | 1215(26.52) | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Type of Cooking fuel** | | | | | | |
| Kerosene or Natural Gas | 3769(20.4) | 415(11.01) | 2838(75.3) | 516(13.69) | | |
| Wood | 8436(45.65) | 790(9.36) | 5690(67.45) | 1956(23.19) | 223.78 | <0.001 |
| Agricultural Crop | 4918(26.61) | 360(7.32) | 3306(67.22) | 1252(25.46) | | |
| Animal Dung and Others | 1356(7.34) | 96(7.08) | 916(67.55) | 344(25.37) | | |
| **Number of living children** | | | | | | |
| No Child | 2025(94.71) | 109(5.1) | 4(0.19) | 2138(100) | | |
| One Child | 0(0) | 4556(99.61) | 18(0.39) | 4574(100) | | |
| Two Child | 0(0) | 6038(97.31) | 167(2.69) | 6205(100) | | **<0.001*** |
| Three Child | 0(0) | 3215(79.32) | 838(20.68) | 4053(100) | | |
| Four Child | 0(0) | 0(0) | 1863(100) | 1863(100) | | |
| Five or more Children | 0(0) | 0(0) | 1294(100) | 1294(100) | | |
| | | | | | | |
| **Body Mass Index** | | | | | | |
| Thin | 2285(11.55) | 315(13.79) | 1453(63.59) | 517(22.63) | | |
| Normal | 11173(56.47) | 1237(11.07) | 7578(67.82) | 2358(21.1) | 158.29 | <0.001 |
| Overweight | 5036(25.45) | 342(6.79) | 3683(73.13) | 1011(20.08) | | |
| Obese | 1290(6.52) | 76(5.89) | 969(75.12) | 245(18.99) | | |
| **Age at first Marriage** | | | | | | |
| <18 | 15166(75.35) | 1192(7.86) | 10315(68.01) | 3659(24.13) | 638.86 | <0.001 |
| 18+ | 4962(24.65) | 834(16.81) | 3603(72.61) | 525(10.58) | | |
| **Contraceptive use and intention** | | | | | | |
| Using modern method | 9854(48.96) | 401(4.07) | 7518(76.29) | 1935(19.64) | | |
| Using traditional method | 1889(9.38) | 77(4.08) | 1229(65.06) | 583(30.86) | 2685.50 | <0.001 |
| Non-user - intends to use later | 5081(25.24) | 1275(25.09) | 3382(66.56) | 424(8.34) | | |
| Does not intend to use | 3304(16.41) | 273(8.26) | 1789(54.15) | 1242(37.59) | | |
| **Son or daughter died** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| No | 17036(84.64) | 2025(11.89) | 12779(75.01) | 2232(13.1) | 4071.07 | <0.001 |
| Yes | 3091(15.36) | 0(0) | 1139(36.85) | 1952(63.15) | | |
| **Access to Media** | | | | | | |
| No | 6847(34.02) | 529(7.73) | 4202(61.37) | 2116(30.9) | 658.04 | <0.001 |
| Yes | 13280(65.98) | 1497(11.27) | 9715(73.16) | 2068(15.57) | | |
| **Sex of household head** | | | | | | |
| Male | 17166(85.29) | 1767(10.29) | 11822(68.87) | 3577(20.84) | 7.78 | 0.020 |
| Female | 2960(14.71) | 258(8.72) | 2096(70.81) | 606(20.47) | | |
| **Ideal number of children** | | | | | | |
| 0-2 Children | 15520(77.61) | 1803(11.62) | 11518(74.21) | 2199(14.17) | | |
| 3-4 Children | 4361(21.81) | 205(4.7) | 2322(53.24) | 1834(42.05) | 1823.94 | <0.001 |
| Over 4 Children | 117(0.59) | 5(4.27) | 33(28.21) | 79(67.52) | | |
| **Currently breastfeeding** | | | | | | |
| No | 2025(12.66) | 10287(64.34) | 3677(23) | 15989(100) | | |
| Yes | 0(0) | 3631(87.75) | 507(12.25) | 4138(100) | | **<0.001*** |
| **Desire more children** | | | | | | |
| Wants within 2 years | 2253(11.87) | 934(41.46) | 1285(57.04) | 34(1.51) | | |
| Wants after 2+ years | 3928(20.69) | 791(20.14) | 3095(78.79) | 42(1.07) | | |
| Wants, unsure timing/Undecided | 577(3.04) | 102(17.68) | 457(79.2) | 18(3.12) | 6110.81 | <0.001 |
| Wants no more | 10256(54.02) | 21(0.2) | 7313(71.3) | 2922(28.49) | | |
| Sterilized (respondent or partner)/Declared infecund | 1970(10.38) | 34(1.73) | 1045(53.05) | 891(45.23) | | |
| **Father's desire for children** | | | | | | |
| Both want same | 14118(79.05) | 1588(11.25) | 10150(71.89) | 2380(16.86) | | |
| Husband wants more | 1939(10.86) | 65(3.35) | 1232(63.54) | 642(33.11) | 565.93 | <0.001 |
| Husband wants fewer | 1089(6.1) | 65(5.97) | 805(73.92) | 219(20.11) | | |
| Don't know | 714(4) | 162(22.69) | 343(48.04) | 209(29.27) | | |

** Fisher’s Exact test were utilized in place of chi-square.*

## Selection of the Best Features

Figure-1 the Boruta algorithm illustrates that we opted to keep 23 variables as a determinant of fertility i.e., Division, Type of place of residence, Wealth index, Maternal educational level, Father's education level, Father's occupation, Mothers' occupation, Number of household members, Type of Toilet Facilities, Sources of Drinking Water , Type of Cooking, Number of living children, Number of living children, Maternal age at first birth, Contraceptive use and intention, Body Mass Index, Son or daughter died, Access to Media, Sex of household head, Ideal number of children, Currently breastfeeding, Desire for more children, Husband's desire for children out of 23 variables are determined most influential factors to predict the fertility in Bangladesh.

## Predicting Fertility using Machine Learning Approaches

The machine learning models namely Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), XGBoost, LightGBM and Neural Network (NN) were utilized to build a model for prediction of fertility in Bangladesh. After training each predictive model with 80% of the data then testing was conducted with the remaining 20% of the dataset.

The Random Forest (RF) model was predicted fertility with accuracy 96.02%. The Random Forest model displayed a consistent proficiency in predicting fertility as evidenced by its macro average recall (91%), precision (89%), F1 score (90%) in addition with weighted average recall (96%), precision (96%), F1 score (96%). The Random Forest model was predicted number of child ever born correctly 271 as no children, 667 as 1 children, 962 as 2 children, 642 as 3 children, 300 as 4 children, 109 as 5 children and 89 as 6 children. It has wrongly predicted 126 child ever born to different class or group (table-3).

The Decision Tree (DT) model was predicted fertility with accuracy 96.02%. The Decision Tree model generally fertility prediction with macro average recall (87%), precision (87%), F1 score (87%) in addition with weighted average recall (94%), precision (94%), F1 score (94%). The Decision Tree (DT) model was predicted number of child ever born correctly 271 as no children, 666 as 1 children, 957 as 2 children, 611 as 3 children, 283 as 4 children, 92 as 5 children and 101 as 6 children. It has wrongly predicted 185 child ever born to different class or group (table-3).

The K-Nearest Neighbors (KNN) model was predicted fertility with accuracy 79.94%. The K-Nearest Neighbors model demonstrated a general ability to predict fertility based on the findings with macro average recall (73%), precision (68%), F1 score (70%) in addition with weighted average recall (79%), precision (80%), F1 score (79%). The K-Nearest Neighbors (KNN) model was predicted number of child ever born correctly 234 as no children, 614 as 1 children, 894 as 2 children, 493 as 3 children, 189 as 4 children, 45 as 5 children and 62 as 6 children. It has wrongly predicted 635 child ever born to different class or group (table-3).

The Logistic Regression (LR) model was predicted fertility with accuracy 95.77%. The Logistic Regression model generally predicted fertility with macro average recall (91%), precision (90%), F1 score (90%) in addition with weighted average recall (96%), precision (96%), F1 score (96%). The Logistic Regression (LR) model was predicted number of child ever born correctly 271 as no children, 659 as 1 children, 954 as 2 children, 643 as 3 children, 300 as 4 children, 113 as 5 children and 92 as 6 children. It has wrongly predicted 134 child ever born to different class or group (table-3).

The Support Vector Machine (SVM) model was predicted fertility with accuracy 96.21%. The Support Vector Machine model exhibited a general effectiveness in predicting fertility as indicated by the macro average recall (93%), precision (89%), F1 score (90%) in addition with weighted average recall (97%), precision (96%), F1 score (96%). The Support Vector Machine (SVM) model was predicted number of child ever born correctly 271 as no children, 668 as 1 children, 963 as 2 children, 643 as 3 children, 299 as 4 children, 131 as 5 children and 71 as 6 children. It has wrongly predicted 120 child ever born to different class or group (table-3).

The XGBoost model was predicted fertility with accuracy 95.55%. The XGBoost model exhibited a general proficiency in predicting fertility based on the findings with macro average recall (90%), precision (89%), F1 score (89%) in addition with weighted average recall (95%), precision (96%), F1 score (95%). The XGBoost model was predicted number of child ever born correctly 271 as no children, 667 as 1 children, 961 as 2 children, 634 as 3 children, 300 as 4 children, 102 as 5 children and 90 as 6 children. It has wrongly predicted 141 child ever born to different class or group (table-3).

The LightGBM model was predicted fertility with accuracy 95.67%. The predictive capabilities of the LightGBM model with macro average recall (90%), precision (89%), F1 score (89%) in addition with weighted average recall (96%), precision (96%), F1 score (96%). The LightGBM model was predicted number of child ever born correctly 271 as no children, 666 as 1 children, 962 as 2 children, 636 as 3 children, 300 as 4 children, 104 as 5 children and 90 as 6 children. It has wrongly predicted 137 child ever born to different class or group (table-3).

The Neural Network (NN) model was predicted fertility with accuracy 94.09%. The predictive performance of the Neural Network model with macro average recall (87%), precision (87%), F1 score (87%) in addition with weighted average recall (94%), precision (94%), F1 score (94%). The Neural Network (NN) model was predicted number of child ever born correctly 271 as no children, 666 as 1 children, 952 as 2 children, 619 as 3 children, 283 as 4 children, 97 as 5 children and 91 as 6 children. It has wrongly predicted 187 child ever born to different class or group (table-3).

**Table-3: Predictive models of performance of Fertility**

| Predictive Models Performances | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **RF** | | **DT** | | **KNN** | | **LR** | |
| **Accuracy** | 96.02% | | 96.02% | | 79.94% | | 95.77% | |
| | Macro Average | Weighted Average | Macro Average | Weighted Average | Macro Average | Weighted Average | Macro Average | Weighted Average |
| **Recall** | 91% | 96% | 87% | 94% | 73% | 79% | 91% | 96% |
| **Precision** | 89% | 96% | 87% | 94% | 68% | 80% | 90% | 96% |
| **F1 score** | 90% | 96% | 87% | 94% | 70% | 79% | 90% | 96% |
| | | | | | | | | |
| | **SVM** | | **XGBoost** | | **LightGBM** | | **NN** | |
| **Accuracy** | 96.21% | | 95.55% | | 95.67% | | 94.09% | |
| | Macro Average | Weighted Average | Macro Average | Weighted Average | Macro Average | Weighted Average | Macro Average | Weighted Average |
| **Recall** | 93% | 97% | 90% | 95% | 90% | 96% | 87% | 94% |
| **Precision** | 89% | 96% | 89% | 96% | 89% | 96% | 87% | 94% |
| **F1 score** | 90% | 96% | 89% | 95% | 89% | 96% | 87% | 94% |

Models with higher recall might be preferred if correctly identifying positive cases (fertility cases in this context) is of utmost importance. Models with higher precision are crucial when reducing false positive cases is a priority. The weighted averages provide insights considering class imbalances, while macro averages treat all classes equally. This analysis suggests that while all models perform reasonably well the choice among them might be based on the specific needs computational complexity and the importance of correctly identifying fertility cases versus minimizing false positives. Figure-2 presents the confusion matrix and figure-3 presents AUROC of all models comparative performances. According to the findings presented in Table-3 the predictive model performance results indicate that the Support Vector Machine (SVM) outperformed other models utilized in this study as the most effective predictor of fertility.

### K-fold cross-validation

K-fold cross-validation was performed for 5-fold, 10-fold, 15-fold, 20-fold and 30-fold the results are organized in Table 4. The SVM model demonstrated superior performance across 5-fold, 10-fold, 15-fold, 20-fold and 30-fold cross-validations consistently achieving higher accuracy scores of 95.90%.

**Table-4: K-fold Cross validation of Machine Learning Models Results**

| Models | Mean Accuracy of K-Fold | | | | |
|---|---|---|---|---|---|
| | **5-Fold** | **10-Fold** | **15-Fold** | **20-Fold** | **30-Fold** |
| **RF** | 0.9496 | 0.9512 | 0.9503 | 0.9496 | 0.9511 |
| **DT** | 0.9348 | 0.9375 | 0.9363 | 0.9354 | 0.9349 |
| **KNN** | 0.5647 | 0.5709 | 0.5707 | 0.5727 | 0.5737 |
| **LR** | 0.9577 | 0.9572 | 0.9579 | 0.9574 | 0.9574 |
| **SVM** | **0.9590** | **0.9590** | **0.9590** | **0.9590** | **0.9590** |
| **XG Boost** | 0.9519 | 0.9534 | 0.9526 | 0.9533 | 0.9537 |
| **Light GBM** | 0.9529 | 0.9535 | 0.9534 | 0.9536 | 0.9545 |
| **NN** | 0.9351 | 0.9380 | 0.9375 | 0.9352 | 0.9373 |

Figure-4 illustrates a graphical depiction of key features identified through the utilization of an SVM classifier. This allows us to distinguish and emphasize the pivotal variables influencing fertility in Bangladesh. The SVM algorithm of Machine learning models aids in discerning the crucial features affecting fertility which include: access to mass media, education level of father, mother educational level, number of household members, Body Mass Index (BMI), number of living children and incidents of sons or daughters died.

# Discussion

This research unveils the discussion on traditional methods and Machine Learning methods and we found that Division, Type of place of residence, Wealth index, Maternal educational level, Father's education level, Father's occupation, Mothers' occupation, Number of household members, Sources of Drinking Water, Maternal age at first birth, Body Mass Index, Contraceptive use and intention, Son or daughter died, Access to Media, Currently breastfeeding, Current marital status, Number of living children, Sex of household head, Ideal number of children, Desire for more children, Husband's desire for children, Type of toilet facilities and type of cooking fuel were the significant factors for predicting fertility in Bangladesh using the machine learning features selection of Boruta algorithm. However, all of the variables were the significant factors only by using conventional chi-square test.

We evaluated the performance of different types of Machine Learning models and it shows that Machine Learning methods predict the factors associated with fertility, the SVM model demonstrated superior performance to predict fertility determinants in Bangladesh. Our best predictive model Support Vector Machine (SVM) identified seven factors i.e., access to mass media, education level of father, mother educational level, number of household members, Body Mass Index (BMI), number of living children and incidents of sons or daughters died as most important as fertility determinant of Bangladesh. Our study shows that Maternal educational level, Father's education level, Access to Media, Son or daughter died are effecting fertility this result is concurrent with the previous research [6]. Our study aligns with the extensive discussions in the literature [22] highlighting the noteworthy association between maternal BMI and fertility. We found that the number of living children increases the probability of expressing a desire for no more children also goes up [26].

In our research found a correlation between fertility and household size, indicating that larger households tend to exhibit higher levels of fertility which is concurrent with the study [27] .

### Strengths and limitations

The most recent country representative BDHS-2017-18 dataset were used in this study. Causal inference is not feasible as this analysis is cross-sectional. The results we examined rely on self-reporting and are thus vulnerable to variation in memory and social desirability. We have used 08 (eight) different types of Machine Learning models: Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, LightGBM and Neural Network (NN) for examining and seeking the best output.

## Conclusion

Machine Learning (ML) models is more reliable over than traditional statistical models for determining of fertility. The machine learning models use training data for constructing a good model and test data for predicting fertility in Bangladesh and then compare the reliability. With the help of our best identified model i.e., Support Vector Machine (SVM) model shows that, Access to mass media, Father's education level, Mother's educational level attained, Number of household members, Body Mass Index (BMI) of mother, Number of living children and Incidents of sons or daughters died are the most important determinant of fertility in Bangladesh.

## Declarations

### Funding

The study does not receive any funding.

### Ethics approval and consent to participate

No ethical approval for this study because this analysis is secondary and research is based on BDHS (2017-18) data. The data were completely anonymous and there is no identifiable information was there and available DHS website.

### Consent for publication

Not applicable.

### Availability of data and materials

Sourced from the publicly available dataset at https://dhsprogram.com/data/available-datasets.cfm the manuscript incorporates secondary data.

### Competing interests

The authors declared no competing interests exists.

### Authors contribution

Conceptualization: Shayla Naznin, Dr. Md Jamal Uddin, Dr. Ahmad Kabir, Data curation: Shayla Naznin, Formal analysis: Shayla Naznin, Methodology: Shayla Naznin, Supervision: Dr. Md Jamal Uddin, Dr. Ahmad Kabir, Visualization: Shayla Naznin, Writing original draft: Shayla Naznin. All authors read and approved the final manuscript.

# References


1. S. Roy and S. M. I. Hossain, "Fertility differential of women in Bangladesh demographic and health survey 2014," *Fertil. Res. Pract.*, vol. 3, no. 1, pp. 3–7, 2017, doi: 10.1186/s40738-017-0043-z.
2. M. F. R. Ahbab, "Factors influencing fertility preference of a developing country during demographic transition : Evidence from Bangladesh," *South East Asia J. Public Heal.*, vol. 4, no. 2, pp. 23–30, 2014.
3. M. Chen1, S. M. A. Haq, K. J. Ahmed, A. H. M. B. Hussain, and M. N. Q. Ahmed, "The link between climate change, food security and fertility: The case of Bangladesh," *PLoS One*, vol. 16, no. 10 October, pp. 1–18, 2021, doi: 10.1371/journal.pone.0258196.
4. M. M. Islam, M. A. Islam, and N. Chakroborty, "Fertility transition in Bangladesh: Understanding the role of the proximate determinants," *J. Biosoc. Sci.*, vol. 36, no. 3, pp. 351–369, 2004, doi: 10.1017/S0021932003006333.
5. S. Islam and M. K. Nesa, "Fertility transition in Bangladesh: The role of education," *Proc. Pakistan Acad. Sci.*, vol. 46, no. 4, pp. 195–201, 2009.
6. R. Adhikari, "Demographic, socio-economic, and cultural factors affecting fertility differentials in Nepal," *BMC Pregnancy Childbirth*, vol. 10, pp. 1–11, 2010, doi: 10.1186/1471-2393-10-19.
7. J. K. Bora, N. Saikia, E. B. Kebede, and W. Lutz, "Revisiting the causes of fertility decline in Bangladesh: the relative importance of female education and family planning programs," *Asian Popul. Stud.*, vol. 19, no. 1, pp. 81–104, 2023, doi: 10.1080/17441730.2022.2028253.
8. I. Haq, "Relationship Between Age at Marriage, Education and Fertility Among Residence of Bangladesh," *Am. J. Soc. Sci. Res.*, vol. 4, no. 2, pp. 33–39, 2018, [Online]. Available: http://www.aiscience.org/journal/ajssrhttp://creativecommons.org/licenses/by/4.0/.
9. Y. Hahn, K. Nuzhat, and H. S. Yang, "The effect of female education on marital matches and child health in Bangladesh," *J. Popul. Econ.*, vol. 31, no. 3, pp. 915–936, 2018, doi: 10.1007/s00148-017-0673-9.
10. A. Kabir, R. Jahan, and S. M. K. Alam, "The Effect of Education on Fertility in Bangladesh," *Int. Q. Community Health Educ.*, vol. 22, no. 3, pp. 199–214, 2003, doi: 10.2190/l0ql-q1uw-b5we-7kd3.
11. M. H. Howlader, H. O. Roshid, S. Kundu, H. R. Halder, S. K. Chanda, and M. A. Rahman, "Determinants associated with high-risk fertility behaviours among reproductive aged women in Bangladesh: a cross-sectional study," *Reprod. Health*, vol. 19, no. 1, pp. 1–10, 2022, doi: 10.1186/s12978-022-01333-w.
12. D. Zelalem, A. Semahegn, G. Tesfaye, and B. Sileshi, "The level and patterns of fertility among women in Kersa Demographic Surveillance and Health Research Center (KDS-HRC) Field site, Kersa District, East Ethiopia," *Fertil. Res. Pract.*, vol. 1, no. 1, pp. 1–8, 2015, doi: 10.1186/s40738-015-0010-5.
13. S. Abedin and D. Arunachalam, "Maternal autonomy and high-risk pregnancy in Bangladesh: The mediating influences of childbearing practices and antenatal care," *BMC Pregnancy Childbirth*, vol. 20, no. 1, pp. 1–8, 2020, doi: 10.1186/s12884-020-03260-9.

14. National Institute of Population Research and Training (NIPORT); ICF International;, “Bangladesh Demographic and Health Survey 2017-18,” 2020.
15. Demographic and Health Survey. The DHS Program. USAID. https://dhsprogram.com/Countries/Country-Main.cfm?ctry_id=1
16. A. Rahman, Z. Hossain, E. Kabir, and R. Rois, “Machine Learning Algorithm for Analysing Infant Mortality in Bangladesh,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13079 LNCS. pp. 205–219, 2021, doi: 10.1007/978-3-030-90885-0_19.
17. E. Mfateneza, P. C. Rutayisire, E. Biracyaza, S. Musafiri, and W. G. Mpabuka, “Application of machine learning methods for predicting infant mortality in Rwanda: analysis of Rwanda demographic health survey 2014–15 dataset,” BMC Pregnancy Childbirth, vol. 22, no. 1, pp. 1–13, 2022, doi: 10.1186/s12884-022-04699-8.
18. R. K. Saroj, P. K. Yadav, R. Singh, and O. N. Chilyabanyama, “Machine Learning Algorithms for understanding the determinants of under-five Mortality,” BioData Min., vol. 15, no. 1, pp. 1–22, 2022, doi: 10.1186/s13040-022-00308-8.
19. F. H. Bitew, S. H. Nyarko, L. Potter, and C. S. Sparks, “Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian Demographic and Health Survey,” Genus, vol. 76, no. 1, 2020, doi: 10.1186/s41118-020-00106-2.
20. J. M. K. Aheto, “Predictive model and determinants of under-five child mortality: Evidence from the 2014 Ghana demographic and health survey,” BMC Public Health, vol. 19, no. 1, pp. 1–10, 2019, doi: 10.1186/s12889-019-6390-4
21. Ayele, Dawit Getnet. “Determinants of fertility in Ethiopia.” *African health sciences vol. 15,2 (2015): 546-51. doi:10.4314/ahs.v15i2.29*
22. Zhu, Lei et al. “Association Between Body Mass Index and Female Infertility in the United States: Data from National Health and Nutrition Examination Survey 2013-2018.” International journal of general medicine vol. 15 1821-1831. 19 Feb. 2022, doi:10.2147/IJGM.S349874
23. Fayemiwo, M. A., and O. Adegbite. "Comparative Analysis of Predictive Models on Fertility Rate Using Machine Learning Algorithms." Adeleke University Journal of Engineering and Technology 2.1 (2019): 85-97.
24. Zhu, Xiaoxia, et al. "Prediction models and associated factors on the fertility behaviors of the floating population in China." Frontiers in Public Health 10 (2022): 977103.
25. Sajid, Mirza Rizwan, Fauzia Maqsood, and Mehwish Rani. "Determinants of fertility: A neural network approach." Middle-East Journal of Scientific Research 8.2 (2011): 440-449.
26. Church, Anna C., et al. "Traditional supports and contemporary disrupters of high fertility desires in sub-Saharan Africa: a scoping review." Reproductive Health 20.1 (2023): 1-16.
27. Bradshaw, Corey JA, et al. "Lower infant mortality, higher household size, and more access to contraception reduce fertility in low-and middle-income nations." Plos one 18.2 (2023): e0280260.

# Figures

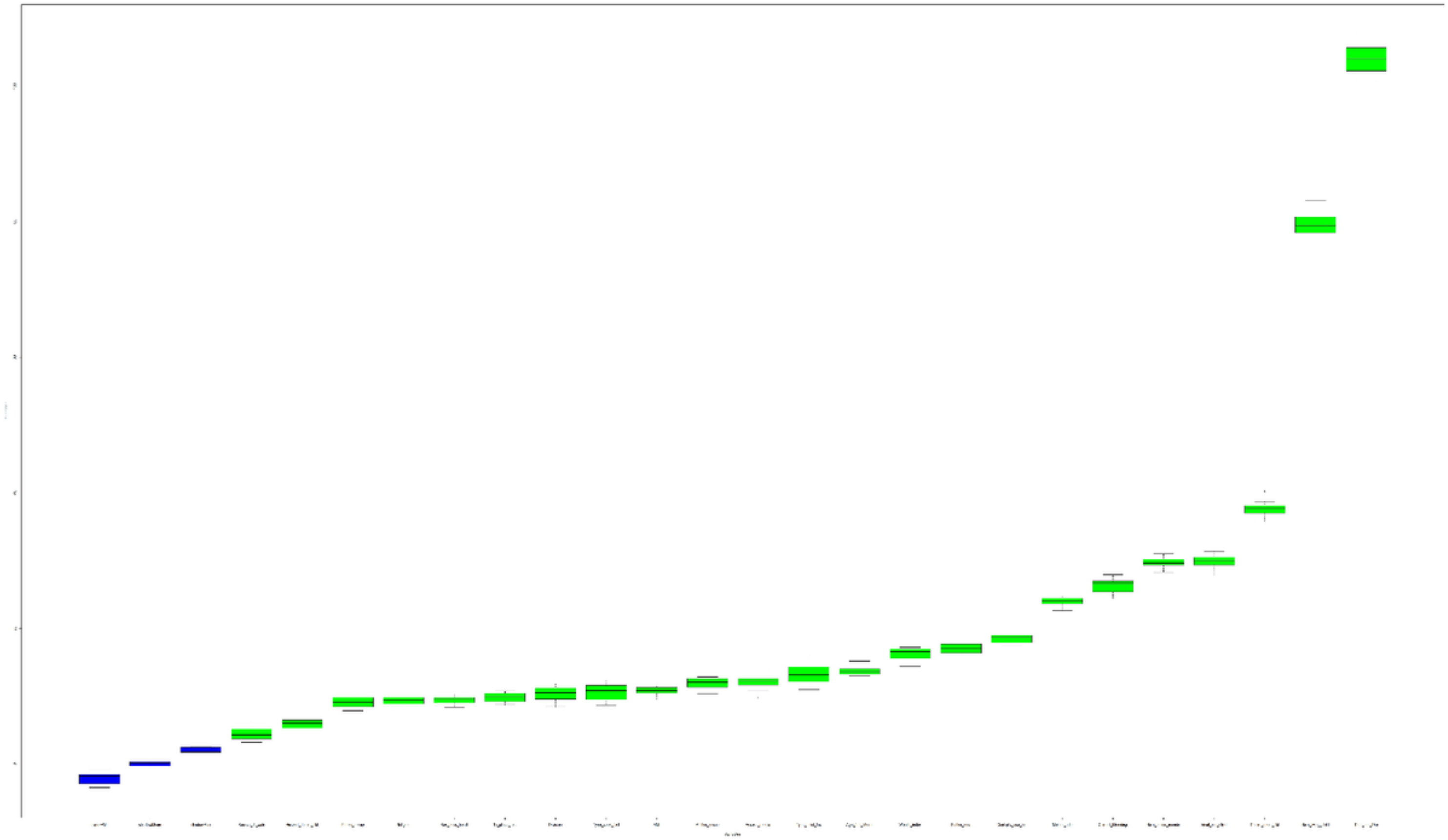
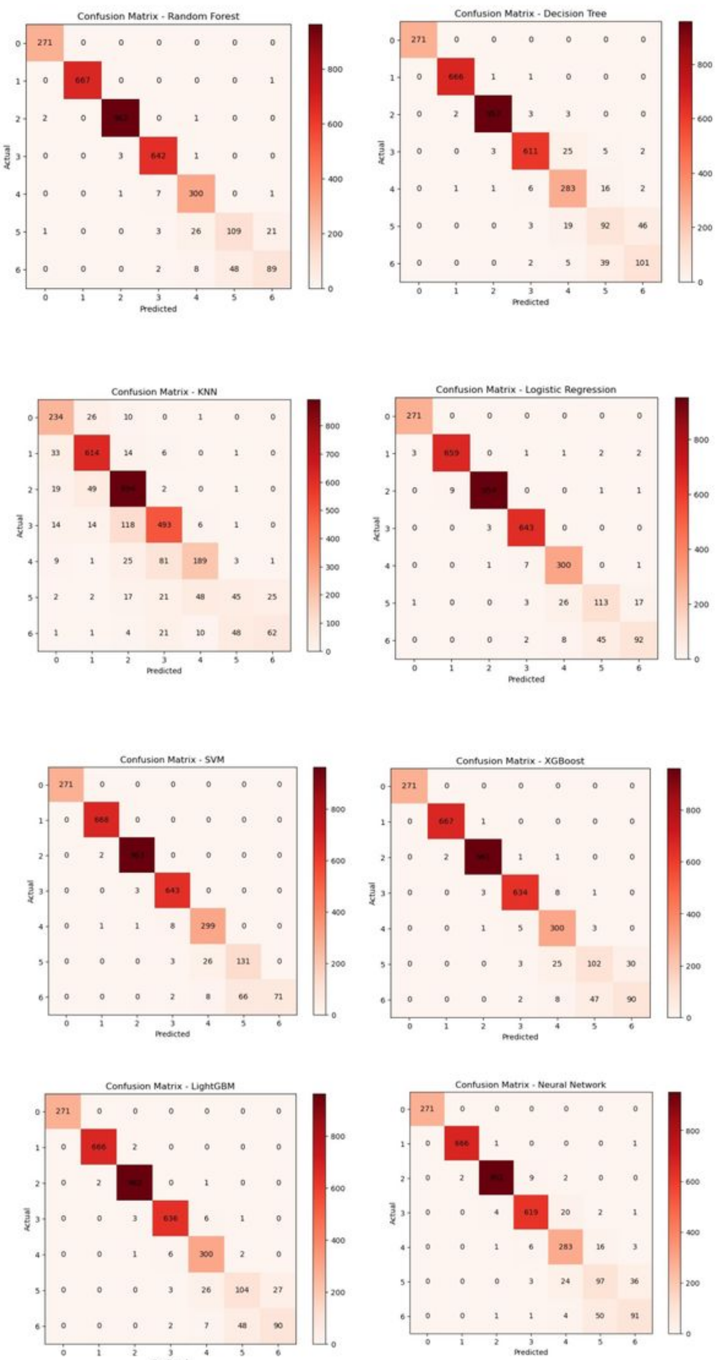


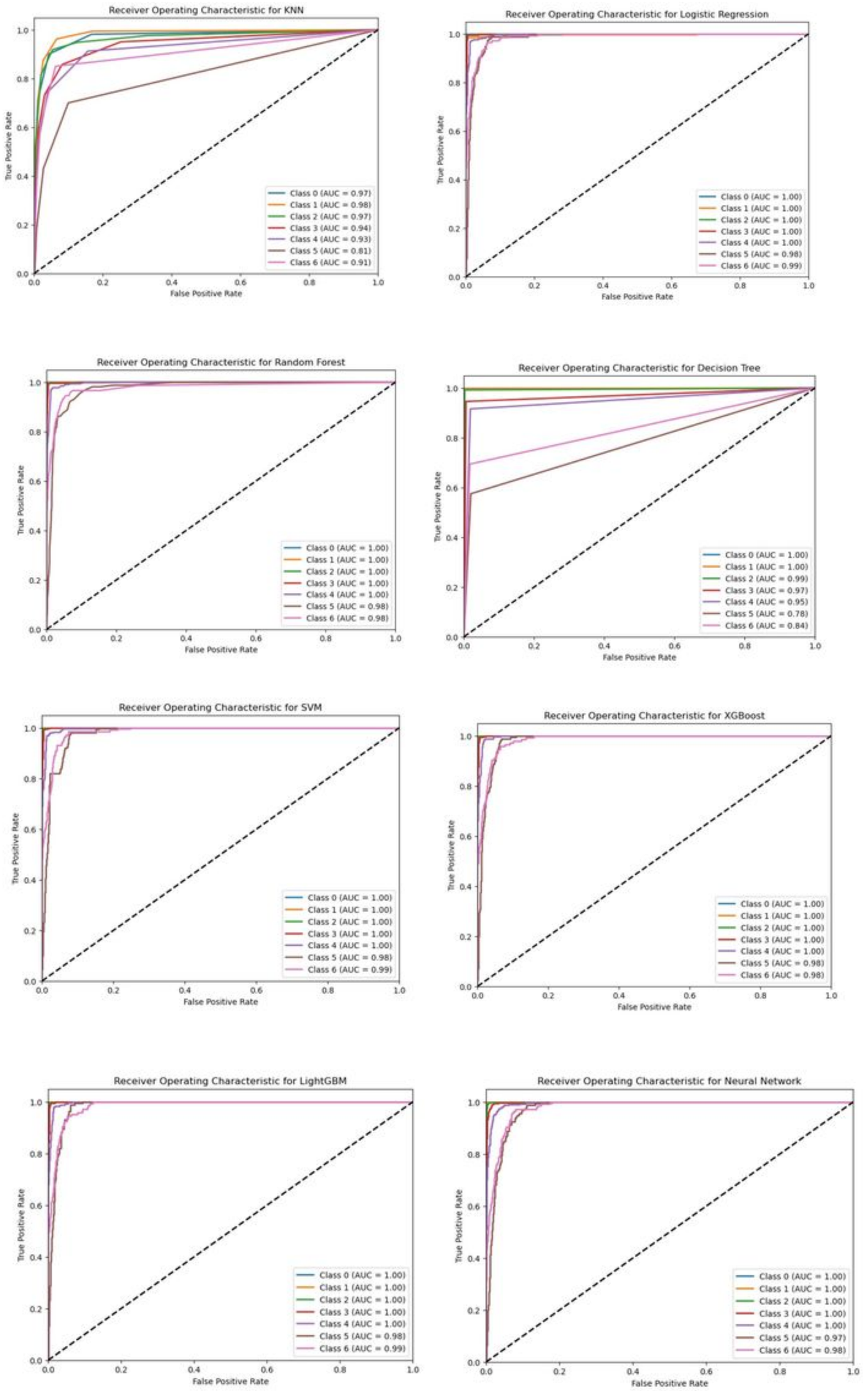
Figure 1

Feature selection using Boruta algorithm




## Figure 2

Confusion matrix of classification results.




## Figure 3

**The ROC curves to predict fertility in Bangladesh using RF, DT, KNN, LR, SVM, XGBoost, LightGBM and NN models.**

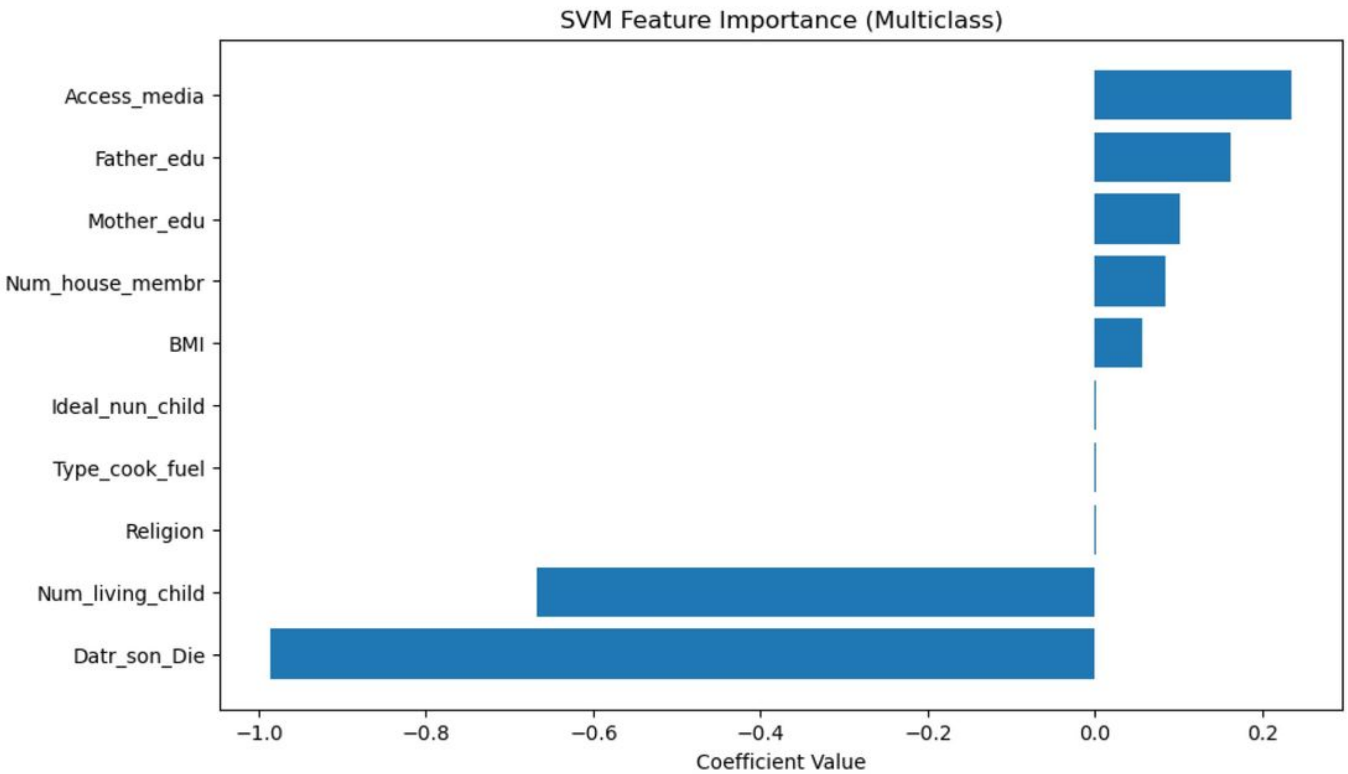



Figure 4

Important features Visualization using ML SVM classifier