National College of Ireland

Socio-Economic and Demographic determinants of Unmet Need for Family Planning Among Indian Women of Reproductive Age

MSc Research Project
Data Analytics

Preeti Sharma
Student ID: 23379031

School of Computing
National College of Ireland

Supervisor: Dr Hicham Rifai

25th January 2026

National College of Ireland

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Preeti Sharma
Student ID:	23379031
Programme:	Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr Hicham Rifai
Submission Due Date:	25th January 2026
Project Title:	Socio-Economic and Demographic determinants of Unmet Need for Family Planning Among Indian Women of Reproductive Age
Word Count:	7604
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Preeti Sharma
Date:	25th January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Socio-Economic and Demographic determinants of Unmet Need for Family Planning Among Indian Women of Reproductive Age

Preeti Sharma
23379031

Abstract

This study investigates the socio-economic and demographic determinants of unmet need for family planning among married women aged 15–49 using the NFHS-5 dataset (2019-21). The dataset contains a total of 724,115 individual records before filtering and approximately 1,400 core variables. A weighted logistic regression model, supported by Elastic Net regularisation and two machine-learning classifiers (Random Forest and XGBoost), was used to analyse associations and validate predictive patterns. Results show that unmet need (prevalence: 10.9%) is strongly shaped by fertility preferences, parity, and information access. Women wanting to delay or avoid further births exhibited the highest risk (OR ≈ 4.09 for "later/unsure"; OR ≈ 3.24 for "no more"). Each additional living child increased unmet need by 24%, while knowledge of modern methods reduced it by 82% (OR ≈ 0.17). Machine-learning models achieved higher discrimination (XGBoost ROC-AUC $= 0.760$), confirming non-linear interaction effects. Overall, unmet need is concentrated among poorer, higher-parity women with limited information exposure, indicating the need for targeted counselling and communication-focused interventions.

Keywords: unmet need, family planning, National Family Health Survey, Chi-Square, logistic regression, machine learning

1 Introduction

Planning a family is one of the main issues in the field of public health, female empowerment, and socio-economic growth. Availability of contemporary contraceptives is associated with the safety of planning and spacing births, decreases maternal and child mortality, enhances poverty alleviation campaigns and promotes gender equality. Even after many decades of policy efforts and family welfare programmes, the unmet need for family planning persists in a large number of women in India.

Unmet need is used to describe women who desire to delay their next birth (spacing) or stop childbearing (limiting) but do not use any contraceptive measures which highlights ongoing service delivery gaps as well as socio-economic disparity, low independence, and cultural obstacles to reproductive choice (Jena et al., 2024; Pasa, N. Singh, Bharali et al., 2024).

Substantial literature demonstrates that an unmet family planning need is determined by a multidimensional factors. Education of women, household wealth, employment and

access to information are some of the socio-economic factors that contribute significantly to contraceptive behaviour. Low-income and low-educated women tend to experience the wrong assumptions of contraceptive side effects, poor mobility, and the lack of access to health services (Alie, Abebe and Negesse, 2022; Oyinlola, Kupoluyi and Adetutu, 2024). The reproductive preference is also influenced by demographic and cultural factors such as age, parity, length of marriage, religion and place of residence. The unmet need is higher in younger women, women with greater parity, and women in rural areas because of structural and socio-cultural limitations worsening the situation of these groups of women into high-risk groups (New, Cahill, Stover et al., 2017).

In convergence with these determinants, methodology in reproductive health research have taken a new shape. Associations between socio-economic conditions and the outcomes of family planning have been traditionally analysed using cross-sectional studies (Maricic, Stojanovic, Pazun et al., 2021). The predictors of contraceptive use have been studied using logistic and multiple regression models, and recent study have used machine learning algorithms like Naive Bayes and Artificial Neural Networks to predict reproductive behaviour more and more often (Devaraj, Gausman, Mishra et al., 2024). Monte Carlo microsimulation has also been used to measure long-term fertility patterns including life-course processes such as union composition, conception, pregnancy loss, childbirth, and fertility decline (S. Singh, Kashyap, Sharma et al., 2023). SEM has contributed to the development of the routes in which socio-demographic and literacy-related variables influence the reproductive intentions.

The socio-economic and demographic environment of India is very diverse with vast discrepancies of states, rural and urban regions, caste and education and cultural environments. Past studies have identified significant disparities in the unmet need by inter-state, with a high concentration of these disparities in the north and center parts of the country, and relative barriers to contraceptive uptake, including fear of side effects, access to counselling or limited access to contraceptive methods(Prasad, Venkatachalam and Z. Singh, 2016). These inequalities should be understood to create equitable, culturally sensitive, and responsive interventions to the lived reality of women in India.

1.1 Objective and Research Question

To what extent do socio-economic and demographic features shape a woman's likelihood of enduring an unmet need for family planning in India?
This paper will analyse how education, household wealth, urban-rural status, information exposure, age, length of marriage, children and religion will determine whether a woman has her need without actually being met. Through such a combination of dimensions, the research aims to find out which of these groups are more vulnerable and to find out why such gaps remain such as they are in the Indian scenario.

The following are the objectives related to this research:

- Re-examining related work on demographic and public-health to explore unmet need within the broad context of reproductive health inequalities.
- Formation and processing of NFHS-5 data: the build a binary unmet-need variable and clean socio-economic and demographic predictor.
- Determinants analysis with survey-weighted logistic regression.

- Use of machine-learning to evaluate predictive performance and confirm the consistency of key predictors using machine-learning models like Elastic Net Logistic Regression, Random Forest and XGBoost.
- Determining the most contributive factors, which socio-economic and demographic factors are regularly at work consistently contributing to unmet need.

The study has three-fold contribution. First, it offers a coherent picture of the most likely women whose need may have been unaddressed and the reasons why some structural barriers still exist. Second, it integrates survey-weighted statistical modelling with optimised machine-learning methods, which provides a more robust analytical method than is typically used in unmet-need studies. Lastly, it produces practical ideas which can be used to strengthen more focused and fairer family-planning in India.

The structure of this report is organised as follows. Section 2 reviews related work and summarises existing studies related to unmet need for family planning. Section 3 explains the methodology followed, including the data source, sampling technique, and the preprocessing steps used to prepare the variables. Section 4 describes the overall design specification of the analytical framework. Section 5 describes how the statistical and machine-learning models were implemented. Section 6 presents the results of these models, followed by a discussion of their consequences. Finally, Section 7 concludes the report and shows potential areas for future work.

2 Related Work

This section covers the literature on unmet need with regard to family planning has identified that there is a very broad scope of socio-economic, demographic, cultural, and programmatic components influencing contraceptive behaviour in different domains. These factors were studied in the past by conventional statistical models, amplified machine-learning methods, and long-term demographic patterns that showed significant regional differences and awareness gaps, as well as service quality gaps. The subsection will show the key themes and methodological techniques that are present in similar studies and serve to form the basis of the integrated analytical framework that will be used in the current research.

2.1 Socio-Economic Determinants

Family planning remains a peripheral aspect of reproductive health care, contributing to women's enablement as well as better maternal and child health, and socio-economic development. There has been an ongoing effort at the policy level where a large number of women from low and middle income countries such as India reported an unmet need for family planning. These are the women who want to postpone or stop pregnancy but are not using any method of contraception. Earlier work has already initiated the fact that this is not a unidimensional phenomenon, rather it emerges out of complex interactions cutting over diverse domains such as socio-economic, demographic, cultural, and informational factors(Jena et al., 2024; Pasa, N. Singh, Bharali et al., 2024).

Comprehensive literature strongly supports the fact that socio-economic attributes are among the major determinants of contraceptive behaviour and unmet needs. The

factors of women's education, household wealth situation, employment status, and access to health information determine the capability and willingness of women to adopt modern methods for contraception (Alie, Abebe and Negesse, 2022). Higher educational levels provide more awareness regarding reproductive health, decision-making sovereignty, and a positive attitude toward family planning; low education typically creates misinterpretations about side effects plus fear of side effects and limited access to health services1(Bangoura, Dioubate, Manet et al., 2021). The economic status of the household further facilitates or limits access to contraceptives and healthcare services since financial constraints on women accessing consultation or modern methods of contraception place limitations(Alie, Abebe and Negesse, 2022).

Also, the lack of reproductive health exposure in the media, talks in the community, or any health literacy program creates deception and further strengthens socio-cultural resistance against contraception(Bangoura, Dioubate, Manet et al., 2021).

2.2 Demographic and Cultural Determinants

Studies have proved that demographic and cultural factors are very important in forming fertility preferences and contraceptive use. The age, duration of marriage, parity, religion, and place of residence were found to be determinants influencing women's feasibility to have an unmet need for family planning. Young women and the early years of marriage most often prioritize fertility, thus resulting in lower contraceptive use. Women with high number of births express in general an increased need to either space or limit births (Prasad, Venkatachalam and Z. Singh, 2016). Religious and cultural norms about what constitutes the ideal family size, fertility expectations, and how acceptable it is to take contraception also influence family planning behaviour.

Paternalistic norms in many, societies deny women the right to reproductive decision-making; thus, they cannot make decisions about contraceptives by themselves. There continues to be a strong rural-urban contrast because unmet needs are higher in rural areas due to limited healthy infrastructure, fewer trained health workers, and cultural barriers against discussing matters related to reproductive health and even choices regarding contraceptives (Sharma, S. Singh and Srivastava, 2021).

2.3 Existing Methodological Approaches

Alongside these factors, there has grown diversity in research methodologies that study reproductive health and fertility behaviour. Though cross-sectional studies have been the most traditional techniques in assessing the association between women's characteristics and unmet need (X. Zhu, Z. Zhu, Gu et al., 2022), other statistical models including logistic regression and multiple linear regression have been used to describe predictors of contraceptive use as well as fertility patterns. More recently, machine learning techniques among which Naive Bayes and Artificial Neural Networks have been used to predict reproductive behaviour within certain populations, show higher predictability strength when evaluated against standard models for complex behaviour (X. Zhu, Z. Zhu, Gu et al., 2022). Monte Carlo micro simulation models essentially provide deeper discoveries by simulating the life course fertility paths of women, including union creation, conception, miscarriage, childbirth, and age-related decline in fertility (Ciganda and Todd, 2021). The synthesis of Structural Equation Modelling results provides an understanding of the trajectories through which women's background features and socio-economic conditions,

together with reproductive health literacy, affect the intentions to practice healthy reproductive behaviours consistently (Maricic, Stojanovic, Pazun et al., 2021). These evolving approaches are indicative of a shift towards more extensive analyses that are also deeply data-driven within reproductive health research.

2.4 Spatial and Regional Variation in India

(Devaraj, Gausman, Mishra et al., 2024) brought forth one of the most holistic long-term perspectives of unmet needs in India by discussing trends from 1993 to 2021 for all 36 states and union territories. Their study validated a steady fall at the national level but pointed a huge inter-state imbalances, with central and northern states always recording higher unmet needs. As indicated by (S. Singh, Kashyap, Sharma et al., 2023), though modern contraceptive usage increased during NFHS 5, inadequacies continue to be highly prevalent-inadequacies for spacing among married women arising out of behavioural, cultural, and method-related obstacles.

Socio-economic determinants are key. As stated by (Rahaman, Rana et al., 2022), women who come from poor households with inadequate education and less contact with the health system in general significantly express more unmet need for spacing contraception. Earlier studies also conveyed similar patterns. Wealth, education, and caste-based disadvantages systematically restrict the avenues through which women can access information and services as argued by Barman (Barman, 2012). More documentation within rural India provided by (Pawar and Solanke, 2020)depicts barriers facing awareness among socio-economically disadvantaged women; hence the constraints are more structural than individual.

Demographic factors determine unmet need. In rural Tamil Nadu, women of higher parity and those in the extended household have a greater expression of unmet need for limiting births, as described by (Prasad, Venkatachalam and Z. Singh, 2016). Younger married women face different problems. The study by (Sharma, S. Singh and Srivastava, 2021) indicated that women in the age group 15–24 years have a high unmet need for spacing because of low independence owing to early marriage and restricted decision-making power. This study also unveiled strong spatial clustering; in fact, unmet need is much higher in the Hindi-belt states.

2.5 Programmatic and Method-Related Barriers

Method-related and programmatic factors are also of key importance. As (Rahaman, Rana et al., 2022) conveyed, even among women desiring contraception, where the major reason is about to express in terms of side effect concern, inadequate mentoring, plus reversible methods not being well available, the unmet need and discontinuation still prevail. This indicates gaps in service quality rather than simple lack of demand.

At the macro level, (New, Cahill, Stover et al., 2017) applied the Family Planning Estimation Tool to analyze trends in contraceptive prevalence and demand across states. Modeling results reveal considerable heterogeneity of met and unmet needs. This emphasizes an explicit inadequacy of a national approach as the sole remedy and emphasizes a necessity for actions on a regional basis.

There is a gap in the literature on the unmet need for family planning as the principal outcome variable in the Indian context using recent nationally representative datasets. Studies that have yet addressed this issue have either taken up specific socio-cultural barri-

ers or mostly managed with contraceptive prevalence, thus not bringing out an integrated perspective regarding multiple socio-economic and demographic influences within a single analytical workflow. India is highly diversified both regionally and socio-culturally, and updated evidence would be obligation to implement a just and targeted family planning intervention. This paper will somewhat try to fill this gap by analysing the socioeconomic and demographic determinants of unmet needs among Indian women of reproductive age.

2.6 Summary

Table 1 presents a consolidated overview of previous work of Section 2, shows the datasets, feature sets, and modelling techniques used across studies.

Table 1: Summary of Key Studies on Unmet Need for Family Planning

Reference	**Dataset / Place**	**Key Variables**	**Model Used**	**Best Finding / Model**
Alie et al. (2022) https://doi.org/10.1186/s40834-022-00168-x	East Africa DHS (12 countries)	Age, education, parity, wealth, media exposure	Multilevel logistic regression	Strong between-country differences
Oyinlola et al. (2024) https://doi.org/10.1371/journal.pone.0306768	Nigeria DHS 2018	Age, marital duration, parity, region, wealth	Multilevel mixed-effects logit	Multilevel model had best fit (AIC/BIC)
Devaraj et al. (2024) https://doi.org/10.1186/s12978-024-01781-6	NFHS 1993–2021 (India)	Fertility preference, parity, SES, media exposure	Survey-weighted logistic regression	Fertility preference & parity strongest predictors
Singh et al. (2023) https://doi.org/10.1038/s41598-023-47191-9	NFHS-5 India	Fertility intention, age, education, caste	Logistic regression	Fertility intention strongest factor
Rahaman et al. (2022) https://doi.org/10.1016/j.cegh.2022.101012	NFHS-4 India	Education, wealth, number of children, religion, region	Multilevel logistic + spatial analysis	Spatial heterogeneity captured (Moran's I)

The literature shows that the lack of family planning need is shaped by complex, multidimensional determinants, including socio-economic, demographic, cultural, and programmatic. Studies have always shown that the education of the women, their wealth in the household, exposure to media, and access to health services are the key point in making decisions that involve contraceptives as well as the lack of these aspects reinforces the falsification, dependency and lack of autonomy. Demographically, age, marital duration, parity, religion, and place of residence are consistently found to be important correlates of contraceptive behaviour with the younger women, rural residence women and those who have higher parity having unevenly high rates of unmet need because of structural and socio-cultural limitations.

Previous work has employed a wide variety of analytical models across methodological strategies, such as cross-sectional modelling, logistic and linear regression, machine-learning methods such as Naive Bayes and ANN, micro simulation, and structural equation modelling, to study predictors of contraceptive use and fertility patterns. These methodologies show a continued changing trend in favour of more data-driven and holistic analyses in reproductive health works.

Spatial and social heterogeneity is also acute as pointed out in the literature of India. National concentrations of unmet needs have decreased, but there is still a wide difference between states, especially in central and northern areas. The unmet need is still found

to be higher among socio-economically vulnerable populations, such as low education of women, lack of access to health systems.

3 METHODOLOGY

In this study, I used a KDD approach (Knowledge Discovery in Databases) to implement the whole analysis. It combines the exploratory analysis, statistical modeling, and machine-learning validation in an organized and systematized manner. The procedures were implemented through conventional procedures with large survey data, such as appropriate weighting and control of variance. The following section concisely describes how the data was collected, how the variables were equipped and the way the models were constructed and tested. Figure 1 shows the overall framework of the analysis of unmet needs used in this study.

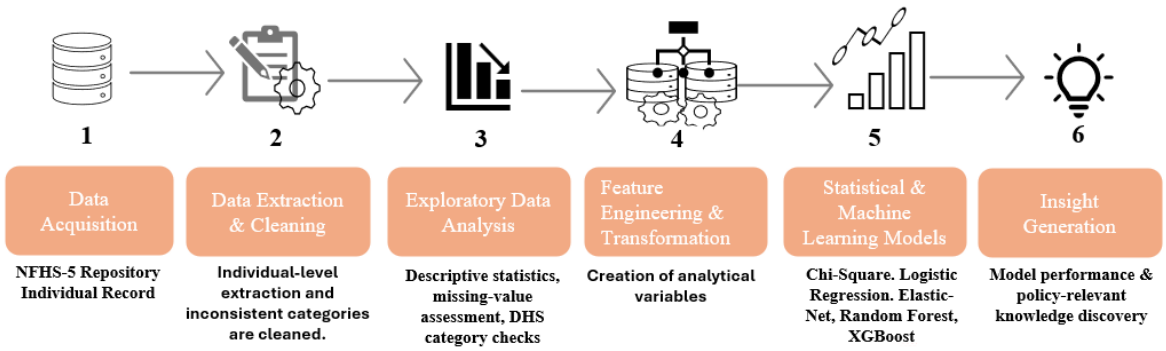

Figure 1: Unmet Need Analysis Framework

3.1 Data Source and Sampling Framework

This study uses nationally representative data from the National Family Health Survey (NFHS-5, 2019–21) and specifically the Individual Recode dataset which includes a total of 724,115 individual records before filtering and approximately 1,400 core variables. Fertility, contraception, socio-economic features as well as maternal health and reproductive behavior of women aged 15–49 years are exhaustively detailed in this dataset for India. The NFHS-5 adopted a two-stage stratified sampling design in its fieldwork for the sample to be nationally representative; hence cluster identifiers (V021), stratification variables (V023), and sample weights (V005) have been used.

The DHS sampling weight (V005) is applied, normalized per DHS guidelines so as to correct for unequal probability of selection and non-response, making the estimates properly reflective of India's female reproductive population rather than the unweighted sample.

Table 2: Description of Variables (NFHS-5 Individual Recode)

Variable / Group	DHS Codes	Type	Description and Analytical Use
Sampling Design and Structure			
Sampling Weight	V005	Continuous	Ensures nationally representative estimates.
PSU / Cluster ID	V021	Categorical	Used for design-based variance adjustment.
Stratification Variable	V023	Categorical	Strata identifier for complex survey design.
Residence (Urban/Rural)	V025	Binary	Infrastructure and access differences.
Cluster	V001	Categorical	Cluster identifier for merges and checks.
Household Number	V002	Categorical	Household-level unique identifier.
Eligibility and Outcome			
Marital Status	V501	Categorical	Unmet need defined for currently married women.
Unmet Need (Outcome)	V626A	Binary	1 = unmet need; 0 = no unmet need.
Socio-Demographic Variables			
Age (Continuous)	V012	Continuous	Exact age in years.
Age Group (5-year)	V013	Categorical	Grouped age (15–49).
Education Level	V106	Categorical	Highest formal education attained.
Years of Education	V133	Continuous	Number of completed schooling years.
Wealth Quintile	V190	Categorical	Household economic status.
Religion	V130	Categorical	Cultural norms influencing fertility attitudes.
Caste / Social Group	V131	Categorical	Structural social stratification.
Fertility Preferences and Reproductive History			
Number of Living Children	V218	Continuous	Parity; determinant of unmet need.
Children Ever Born	V201	Continuous	Total births to respondent.
Fertility Preference (3-category)	V605	Categorical	Desire for more children (Soon / Later / No more).
Age at First Birth	V511	Continuous	Indicator of early childbearing.
Knowledge and Media Exposure			
General Media Exposure (3-item index)	V157-V159	Composite	Exposure to newspaper, radio, TV; constructed as score.
FP Message Exposure (Mass Media)	V384A-V384C	Binary / Composite	Exposure to family planning messages via TV, radio, or media.
Heard FP Message (Any Source)	V364	Binary	General exposure to FP information.
Knowledge of Modern Methods	V301	Binary	Knows at least one modern contraceptive method.
Autonomy and Healthcare Barriers			
Healthcare Access Barriers (8-item index)	V467B-V467I	Composite	Barriers: permission, money, distance, going alone, no female provider, no provider, distrust, long waiting time.

3.2 Data Cleaning and Pre-processing

3.2.1 Data Handling and Cleaning

From the NFHS Individual Recode file, a restricted set of variables was selected covering survey design (V005, V021, V023, V001, V002, V024, V025), the unmet-need variable (V626A), marital status (V501), fertility history (V201, V218, V511), fertility preferences (V605), socio-economic position (V013, V106, V133, V190), media and family-planning information exposure (V157-V159, V301, V384A-V384C, V364), health-system barriers (V467B-V467I) and social background (V130, V131).

A standard set of procedures was used to clean all the variables that had been selected. Design variables were normalised and typed, i.e., sampling weight V005 was divided by 1,000,000 to get a probability weight and that using integer variables: cluster and stratum labels (V021, V023, V001, V002) were put as integer variables to be used in survey weighted analysis.

3.3 Exploratory Data Analysis

In this section some key insights of demographic and socio-economic patterns are presented.

Figure 2 shows a clear concentration of in-union women between ages 22 and 35, where the density is highest. The smoothed density curve reinforces this pattern and highlights a steady decline in representation after the mid-30s. Overall, the graph reflects the typical age structure of married women captured in NFHS-5.

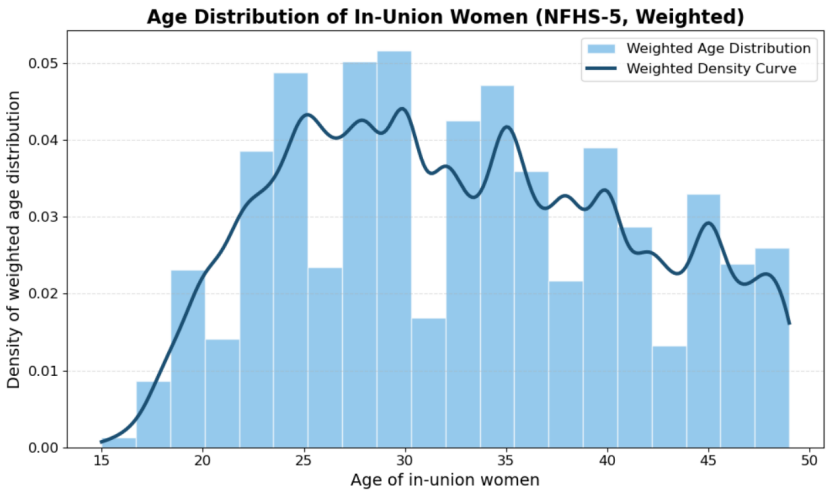

Figure 2: Weighted age distribution of in-union women

In Figure 3, the unmet need for family planning is highest among younger women aged 15–24, with prevalence progressively declining across older age groups. Women in their late twenties and early thirties still exhibit noteworthy unmet need, but levels fall sharply after age 35. This pattern shows both changing fertility intentions and increased contraceptive uptake with age.

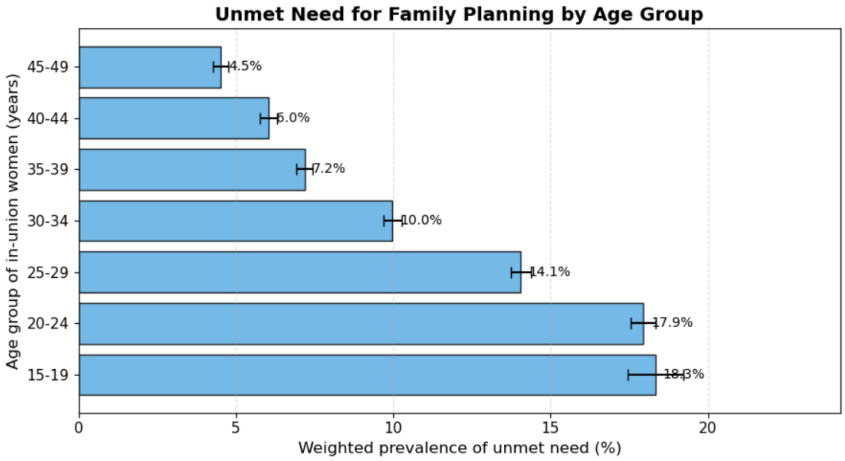

Figure 3: Unmet need for family planning by age group

The unmet need for family planning increases uniformly as the number of living children rises (Figure 4). Women with one or two children have relatively low unmet need, whereas those with four or more children exhibit substantially higher levels. This pattern suggests growing contraceptive demand and spacing/limiting pressures as family size increases.

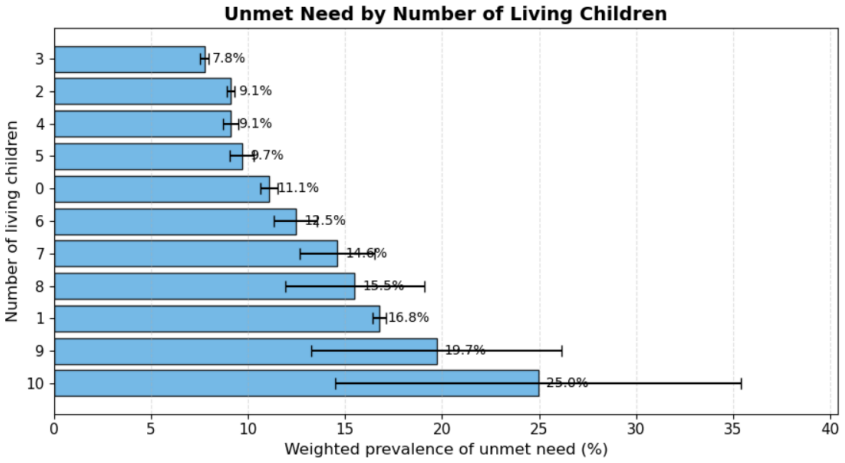

Figure 4: Unmet need for family planning by parity

Figure 5 highlights that poorer women have a higher unmet need for family planning, and this need decreases as wealth increases. Women who face healthcare-access barriers regularly show slightly higher unmet need than those who do not. The biggest difference between the two groups appears among the poorest women.

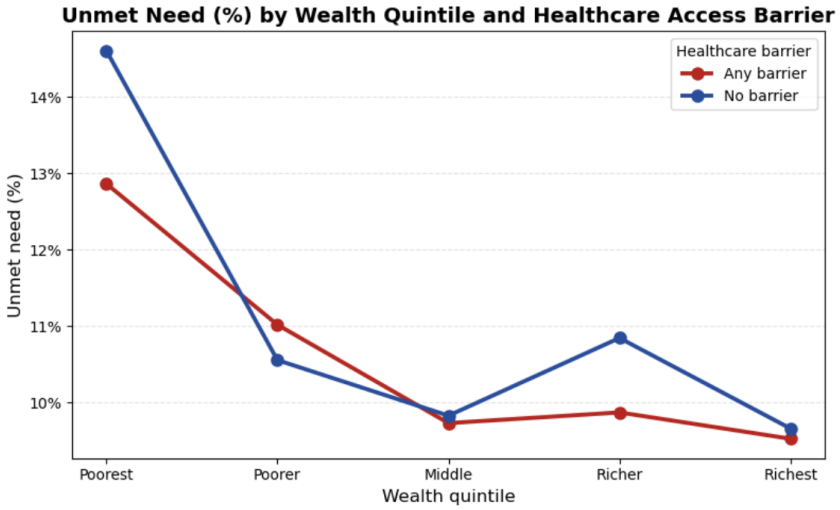

Figure 5: Unmet needs across wealth group

Unmet need for family planning in Figure 6 increases as women's education increases, and rural women consistently have higher unmet need than urban women at every education level. The difference becomes especially large among women with higher education.

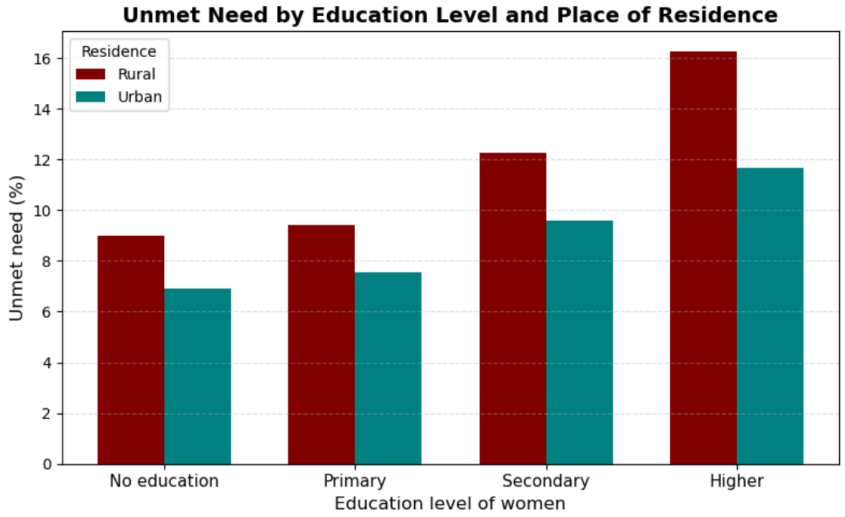

Figure 6: Unmet need across education groups for rural and urban women

Figure 7 shows that unmet need not met by family planning is always high among women who had low exposure to information, especially women who had no contacts with the mass-media or were not exposed to family planning messages. The trend also reveals that the barriers to healthcare access, which even in case of minor ones are present, are connected with a marginally higher unmet need. In general, the figure shows that the absence of information and access together leads to a lack of reproductive needs.

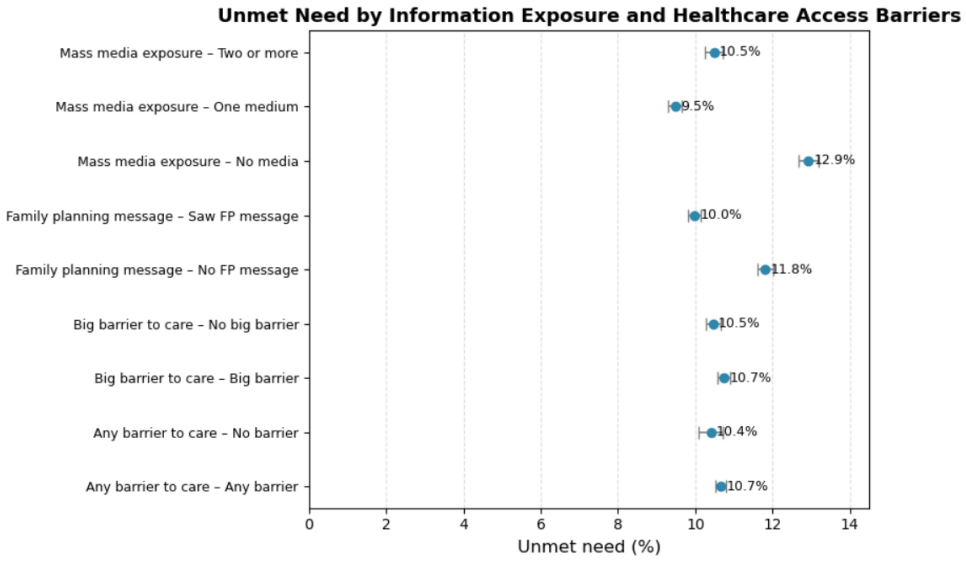

Figure 7: Unmet need by exposure and barriers.

3.4 Feature Engineering

Feature engineering involved transforming raw DHS variables into analytically insightful predictors suitable for statistical and machine-learning models. Multiple groups of engineered features were created which reflecting socio-economic, demographic, reproductive, informational, and healthcare-access dimensions.

3.4.1 Recoding of Socio-Economic and Demographic Variables

Several DHS categorical variables were converted into ordered or labelled factors to preserve their interpretive structure.

- Place of residence (V025) was recoded into a binary Urban/Rural variable.
- Age group (V013) was transformed into ordered 5-year age groups.
- Education (V106) was recoded into a four-level order (no education, primary, secondary, higher).
- Wealth index (V190) was saved as an ordered quintile variable.
- Religion (V130) and caste (V131) were converted from numeric DHS codes into labelled categorical variables.

3.4.2 Reproductive and Fertility-Preference Feature Construction

Raw fertility variables from NFHS-5 were cleaned and augmented into structured predictors.

- Number of living children (V218) was cleaned and transformed into a capped parity variable to reduce scarcity from extreme values.
- Age at first birth (V511) was converted to a numeric variable and implausible values (≥ 95) were reassigned as missing.

- Fertility intention (V605) was recoded into multi-level categories obtaining whether women want a child soon, later/unsure, or no more.

3.4.3 Media Exposure and Family-Planning Knowledge Features

Multiple composite informational features were derived from DHS communication indicators.

- Media exposure (V157–V159) was converted into binary indicators.
- Family-planning knowledge (V301) was recoded into an ordered knowledge-level factor and binary indicators representing any knowledge, modern method knowledge, conventional only knowledge, and no knowledge.
- Exposure to FP messages (V384A-C) was aggregated into a count variable and a binary indicator for exposure over at least one channel.
- These features captured women's access to information, one of the strongest predictors of contraceptive behaviour.

3.4.4 Healthcare-Access Barrier Engineering

Healthcare barrier variables (V467B-V467I) were decomposed and accumulated to quantify structural constraints.

- Each barrier label was split into two engineered indicators: "big problem" and "any problem."

3.4.5 Outcome Variable Engineering

The unmet-need variable (V626A) was cleaned and transformed:

- Outcome indicating unmet need for spacing/limiting versus no unmet need.
- A five-category unmet-need classification differentiating spacing, limiting, no need, current use, and infecund/menopausal status.

This created a consistent and analytical target variable for both statistical and predictive models.

4 Design Specification

The overall structure of the analytical system that has been produced during the present research is structured into two major layers (Figure 8). Both layers have a specific role in processing, modelling, and interpretation of unmet need for family planning by using the NFHS-5 Individual Recode dataset. Below is the description of the layers.

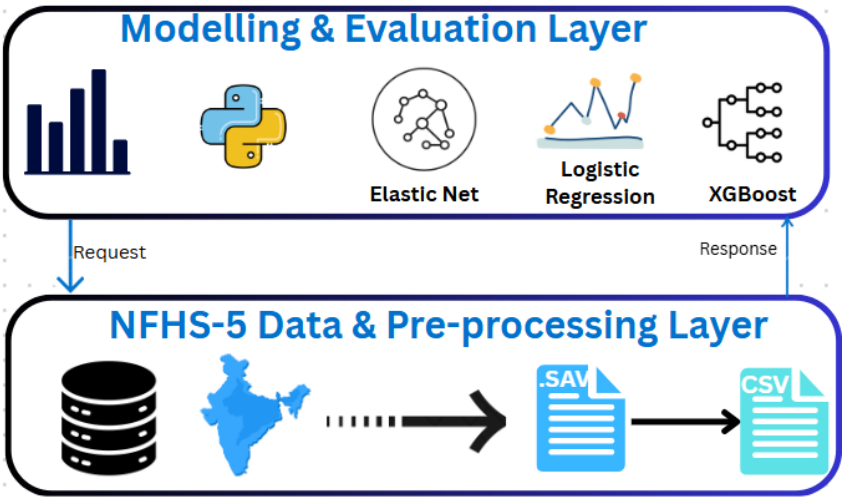

Figure 8: Design Specification

4.1 Database Layer

The first layer consists of uploading the dataset and applying the different pre-processing steps. Feature engineering is performed at this phase done here, this includes creating interaction terms, composite indices like empowerment and healthcare-accessibility scores and non-linear transformations of variables with curvilinear demographic relationships. In prior studies on NFHS data based approach points out the importance of making better socio-demographic and accessibility indices when analysing unmet need (Pawar and Solanke, 2020).

4.2 Application Layer

The second stage is concerned with the statistics modelling and machine-learning based predictive analysis. This layer includes models such as penalised logistic regression, Understandable Boosting Machines, and gradient-boosted decision tree models. Model performance includes AUC, accuracy, F1-score, and subgroup impartiality tests based on caste, wealth, and residence, the approach aligned with NFHS studies that use predictive models and comparative performance metrics to analyse contraceptive behaviour (Rahaman, R. Singh and Chouhan, 2022).

5 Implementation

5.1 Bivariate Association Analysis: Chi-Square Rationale

Prior to development of the multivariate logistic regression model, weighted Chi-square test was used to estimate the bivariate relation among unmet need for family planning and each categorical variable to understand the correlation between different variables and unmet need.

5.2 Statistical Modelling Approach

5.2.1 Classical Logistic Regression

The analytical framework in estimating the determinants of unmet need for family planning was a survey-weighted logistic regression, which was done with the Generalised Linear Model workflow. This classical methodology is popular in the study of demographic and public-health data since the resulting coefficients are comprehensible, statistically valid inference is obtained and it supports design based weighting in line with DHS/NFHS sampling methodology.

The model was estimated using individual survey weights and heteroskedasticity strong HC1 standard errors to account for the complex survey design. After estimation, the analysis produced coefficient estimates, odds ratios with 95% confidence intervals, and average marginal effects, which summarise the associations with unmet need. Because the objective of the analysis was inference rather than prediction, the full analytical sample was preserved and no train–test split was performed, consistent with DHS/NFHS analytical standards.

Model performance was evaluated using the area under the ROC curve, area under the Precision - Recall curve, and threshold-based evaluation at a policy-relevant cutoff $\tau = 0.145$.

5.3 Machine-Learning Logistic Regression (Elastic Net)

To complement the classical GLM and assess predictive robustness, a machine-learning variant of logistic regression was implemented with Elastic Net regularization. This model applies a blend of L1 (lasso) and L2 (ridge) penalties and is well suited for high dimensional feature spaces resulting from one-hot encoding of categorical variables.

Training of the model was done using 80/20 stratified split. In comparison to the survey-weighted GLM, no sampling weights were used in this predictive scenario, because machine-learning models do not seek to do population inference but to maximize the performance of generalization.

The held-out test set was used to assess predictive performance using ROC-AUC, PR-AUC and threshold-based measures at $\tau = 0.145$. The performance of the tuned Elastic Net model was almost equivalent to the classical GLM, and it proved that in terms of the paradigm of inferential and predictive modeling, the factors of unmet need are consistent.

5.4 Justification for Dual Model Framework

The machine-learning logistic regression and the classical GLM apply to different methodological purposes. GLM enables inference, estimation of effect size, and interpretation of the policy under complicated survey design. Elastic Net is a regularized predictive model with cross-validation that is used to estimate how strong the results are to varying modeling assumptions. The fact that the results of the two models converged is a confirmation of the trustworthiness and consistency of the experiential determinations.

5.5 Random Forest

Random Forest is an ensemble machine which utilizes numerous decision trees to make a more stable accurate prediction. It builds on the concept of bagging whereby each of the trees is trained on a random sample of the data which support the model in reducing variance and prevent overfitting. RF was applied in this study based on Scikit Learn's Random Forest Classifier on the same set of socio-demographic and reproductive predictors that were used in the logistic regression model. As tree algorithms have the ability to deal with category codes in their naturally happening form, I transformed all categorical variables into numerical category labels.

The dataset was split into training (80%) and testing (20%) using stratification to preserve the proportion of unmet need cases. DHS sampling weights were included by use of sample_weight argument to maintain the representativeness of the survey. The model evaluated at a fixed threshold of 0.145 to maintain comparability with logistic regression.

5.6 Extreme Gradient Boosting (XGBoost)

XGBoost is also an ensemble tree based method, but it employs the boosting method in place of the bagging method. In boosting, every new tree attempts to fix the errors committed by the earlier trees making the process of learning more focused and efficient. XGBoost is also reputed to be very accurate and perform well especially on tabular datasets. It was implemented using the XGB Classifier, again with the same predictors and weights of the surveys to RF and logistic regression. All nominal variables were coded as numbers, in categories, and the 80/20 stratified division was made once again.

The optimal combination of parameters was chosen and the tuned model was tested on the test set, further testing on the threshold of 0.145 was done so that the model is matching to the other models.

6 Evaluation

6.1 Logistic Regression Model Evaluation

The logistic regression equation was formulated to investigate the influence of the major socio-economic and demographic factors that determine the probability of a woman to have an unmet need on family planning in India. In the model, 452,695 in-union women were used, which is based on the NFHS 5 results that included the entire category of variables in the research question age, education, wealth status, caste, residence, fertility history, media exposure, knowledge of modern methods, and fertility desires. DHS sampling weights and strong HC1 standard errors had been used to have nationally representative and statistically valid estimates.

The model demonstrates high level of distinctive power as shown by ROC AUC of 0.741 indicating that the predictors together have a discrete power in differentiating between needy and non-needy women. The PR AUC of 0.244 is in line with the low national prevalence of unmet need (10.9%) yet shows significant precision in determining vulnerable women. These measures validate the assumption of the research question as the socio economic and demographic factors do offer significant predictive value regarding unmet need.

Using a probability threshold of 0.145, which was policy-consistent, the model had an accuracy = 0.768, recall = 0.543, and precision = 0.246. The output of the recall is notably significant to this study: it is possible to find more than half of the women who actually have unmet need with the help of this model, and this is its usefulness in understanding the vulnerable groups. The confusion matrix also indicates that the model shows 26,911 true cases of unmet need, which is a high level of sensitivity in a situation of public-health where under identification may result in involvement being missed.

On the whole, the analysis shows that socio-economic and demographic factors provide a meaningful explanation of unmet need for family planning. The desire of fertility, age, education, children, exposure to the media and awareness of the modern techniques are some of the factors that contribute considerably to the indicated likelihoods in the direction of the demographic theory. Logistic regression is thus a powerful predictive model and a platform that can be interpreted which directly responds to the research question by quantifying the contribution of each determinant to the risk of unmet need in the Indian women.

6.2 Elastic Net Logistic Regression Evaluation

The Elastic Net penalty was used as a regularised logistic regression model to deal with the multicollinearity existing between socio-economic predictors and to increase model generalisation. One-hot encoded categorical variables in a pipeline and numeric and binary variables were relayed through. Stratified GridSearchCV (with DHS sampling weights) with hyperparameters (regularisation strength C and mixing ratio l1_ratio) was used to select the best hyper parameters to maintain national representativeness.

The optimal model was C = 0.1 and l1_ratio = 1.0, which implies that a high L1 penalty (Lasso) yielded the most stability results by reducing irrelevant predictors. The tuned model recorded ROC-AUC = 0.741 and PR-AUC = 0.245 and equal performance at the policy threshold (0.145). This establishes that the Elastic Net regression is effective in identifying the most important socio-economic and demographic variables that are related with the need unmet to plan a family.

6.3 Random Forest Model Evaluation

To evaluate the existence of non-linear or interaction effect on socio-economic and demographic patterns connecting to unmet need to have family planning, the hyperparameter-tuned Random Forest model was used as a non-parametric model standard. The best configuration found after tuning (with 5-fold Stratified GridSearchCV) was 300 trees with max depth=15, min samples split=50, min samples leaf=10, and max features= sqrt, which means the regulated and reasonably deep trees were the best at generalising on this dataset.

The overall predictive performance of the tuned Random Forest was 0.757 ROC-AUC and 0.254 PR-AUC, respectively, which was better than the classical weighted logistic regression. However, the model reached accuracy of 0.748, precision 0.239, and recall 0.595 with an F1 score of 0.341 at a policy-relevant threshold of 0.145. These findings demonstrate that Random Forest is more sensitive when it comes to deciding the women who are in need, but it produces a somewhat higher false positives than logistic regression. According to the confusion matrix, the model accurately detected 5899 cases of true positives, which means that it is better suited to reveal vulnerable groups.

In sum, the Random Forest model supports the idea that there is no reason to assume that socio-economic and demographic factors that contribute to the unmet need are strictly linear and can be freely modelled. However, since the motive of the study is interpretability and not pure prediction, the logistic regression is the main analytical tool, and the Random Forest is used as a reinforcement of the resilience test.

6.4 XGBoost Model Evaluation

The XGBoost classifier hyperparameterized was used to investigate non-linear, and interaction based relationships between socio-economic and demographic predictors of unmet need of family planning. Using a 5-fold Stratified GridSearchCV, the optimum parameter combination was 400 trees, max_depth = 4, learning_rate = 0.05, subsample = 0.8, colsample_bytree= 0.7. This indicates that the most reliable performance is given by moderately deep trees using restrained feature sampling and a learning rate that is conventional on a large and imbalanced dataset of DHS.

Table 3: Comparison of Model Performance Metrics for Predicting Unmet Need

Model	**ROC-AUC**	**PR-AUC**	**Accuracy**	**Recall**
Baseline Logit	0.741	0.244	0.768	0.543
Tuned Elastic Net Logit	0.741	0.245	0.768	0.541
Tuned Random Forest	0.757	0.254	0.748	0.595
Tuned XGBoost	0.760	0.260	0.751	0.597

Table 4: Performance metrics and confusion matrix components at threshold = 0.145

Model	**Precision**	**F1 Score**	**TN / FP**	**FN / TP**
Baseline Logit	0.246	0.338	320,547 / 82,581	22,656 / 26,911
Elastic Net Logit	0.246	0.338	64,150 / 16,476	4,550 / 5,363
Random Forest	0.239	0.341	61,863 / 18,763	4,014 / 5,899
XGBoost	0.242	0.344	62,112 / 18,514	3,998 / 5,915

The model shows a ROC-AUC of 0.760 and a PR-AUC of 0.260, and implies that XGBoost is more successful at capturing complex, non-linear patterns in unmet need in comparison to logistic regression or Random Forest. The model had a 0.751, 0.242, and 0.597 accuracy, precision, and recall, respectively, at the threshold of 0.145, and this gave a F1 score of 0.344. The fact that the model identifies proportion of true unmet need cases (5,915 true positives) is greater than in the linear and bagging-based models indicates that the model is more effective in correcting patterns which could be overlooked in other models.

Nevertheless, like Random Forest, the coefficients of XGBoost do not hold out explanatory contribution to the study but a predictive one. The model is a robustness check that validates that unmet need does not only depend on additive socio-demographic factors but also on the interaction of parity and fertility preferences or education and media exposure which are according to the theory of demographic studies.

6.5 Discussion

The results of this paper have clearly indicated that the need among the Indian women that is not being met with family planning is not equally distributed but is highly influ-

enced by a combination of socio-economic, demographic and information related factors. In all four modelling methods, namely, the Weighted Logistic Regression, Elastic Net Logistic Regression, the Random Forest, and the XGBoost, common determinants were observed, thereby assisting the structural significance of the given predictors.

The most significant determinant of all the models was fertility preference. Quantitatively, women, who desired to not have a child, were more than four times more likely to have unmet need (OR $\approx$ 4.09), and women, who did not desire any more children, were greater than three times more likely to do so (OR $\approx$ 3.24). This is in line with the previous demographic literature in India with fertility intention indicating to be one of the most reliable predictors of contraceptive behaviour (Devaraj, Gausman, Mishra et al., 2024; S. Singh, Kashyap, Sharma et al., 2023). Other studies which investigate NFHS-5 also affirm that women who desire not to have a pregnancy usually encounter significant obstacles to seeking appropriate methods.

Parity and the number of live children was also a good predictor, where an increment in the number of children raised the odds of unmet need (OR $\approx$ 1.24). Previous studies have also reported high need among women with two or more children which is not met because of spacing/limiting mismatches and gaps in service quality (Pawar and Solanke, 2020; Prasad, Venkatachalam and Z. Singh, 2016). These results indicate larger demographic trends reported in the Indian and South Asian fertility researches.

There were socio-economic inequalities. The poorer households with women had been consistently higher in their unmet need despite patterns of increased inequality related analysis in the work of many inequality based analyses (Jena et al., 2024). Education showed some subtle influence with increased educational attainment corresponding to small difference in unmet need in the regression models. This is reflective of findings by recent research on women education and contraceptive behaviour (Pasa, N. Singh, Bharali et al., 2024), which refers to the possibility of educated women better expressing fertility objectives, which is nonetheless limited by some form of social negotiation, stereotyping, or disappointment with methods.

Variables pertaining to information proved to be strong protective factors. The awareness of simultaneous practices decreased the unmet need by 82 percent (OR $\approx$ 0.17), and awareness of family planning messages by 18 percent (OR $\approx$ 0.83). The following results are well-supported by previous studies that show the influence of the media, FP messages, and reproductive literacy on enhancing contraceptive adoption (Bangoura, Dioubate, Manet et al., 2021; Maricic, Stojanovic, Pazun et al., 2021; Rahaman, Rana et al., 2022). The correspondence of these findings between classical and machine learning models provides support to the centrality of access to communication in the resolution of unmet need.

In general, the findings suggest that women who: (i) are in the middle or later reproductive stage, (ii) have two or more children, (iii) live in lower-wealth families, (iv) are strongly motivated to have no more births, but (v) do not have sufficient information or decisional autonomy to adopt contraception are most likely to have an unmet need. These trends can be attributed to the well-known demographic studies and show that parity-specific consultation and better access to modern techniques should be used.

7 Conclusion and Future Work

This study aimed at establishing socio-economic and demographic predictors of need unmet by the family planning of reproductive age Indian women using NFHS 5 data. In the four modelling models, which are: the Weighted Logistic Regression, Elastic Net Logistic Regression, the random forest and the XGBoost, the results consistently reveal that unmet need has a structured effect of combination of fertility intentions, demographic pressures, socio-economic inequality, and access to information.

The strongest and most consistent determiner of all the models was fertility preference. The higher the predicted likelihood of unmet need was, indicating that there was a continued discrepancy between reproductive intentions and contraceptive services access, were women who wanted to either delay or stop bearing children. Also very strong positive relations were found between parity and the number of living children indicating that unmet need is more focused as women progress through their reproductive life course. The socio-economic limitation also increased vulnerability. The unmet need was significantly higher among poorer women, and the mixed pattern with education can also be explained by the availability of the methods.

Predictors that were information related exhibited strong protective effects. The presence of knowledge about the current contraception procedures decreased the unmet need by about 82% and exposure to family planning messages decreased it by about 18%. These results support the significance of communication channels and reproductive literacy. The machine learning models also showed that joint interactions, these include urban residence and high parity, caste and media exposure, among others, have an effect on unmet need, which indicates that the effect of disadvantage is not through independent elements, but through multi-layered socio-demographic processes.

All these data points to the fact that unmet need is greatest in women who do not want to get pregnant but do not have information, autonomy, or access to services necessary to take action on their intentions, mainly in poorer and high-parity households. The fact that the results of statistical and machine learning models are consolidating supports the validity of this conclusion and is consistent with previous demographic studies in India.

7.1 Future Work

Current analysis of this study gives good explanatory and predictive explanations, there are still some directions where research can be undertaken in the future. To begin with, it might be useful to incorporate geospatial or district-level indicators of the health system (e.g., facility density, ASHA outreach, stock availability) to reveal spatial heterogeneity that is not determined by the existing models. Second, causal inference designs including marginal structural models or propensity score designs might be used to differentiate between behavioural drivers and structural hindrances. Third, longitudinal or repeated cross-sectional NFHS wave modelling can help identify whether the factors of unmet need are changing with time. Fourth, high-risk group qualitative studies would provide more detailed information on the power of negotiation, cultural dimensions, and decision-making processes at the household level. Lastly, example simulation models with a policy focus such as: what would be the impact of raising the media exposure or expanding the range of potential methods might be useful in facilitating specific intervention.

All in all, this paper shows that a lack of need in family planning in India is informed by overlapping demographic strains, socio-economic disparities and lack of reproductive

information. These multi-layered constraints will be fundamental concerns in addressing the unmet need and empowering reproductive autonomy of women in all parts of India.

8 Acknowledgement

I express my gratitude to Professor Hicham Rifai for his valuable guidance which helped me in my analysis. And I am also thankful to the National College of Ireland for providing the essential resources.

References

Alie, M.S., G.F. Abebe and Y. Negesse (2022). "Magnitude and determinants of unmet need for family planning among reproductive age women in East Africa: multilevel analysis of DHS data". In: *Contraception and Reproductive Medicine* 7.1. DOI: 10.1186/s40834-022-00168-x. URL: https://doi.org/10.1186/s40834-022-00168-x.

Bangoura, C., N. Dioubate, H. Manet et al. (2021). "Experiences, preferences and needs of adolescents and urban youth in Conakry, Guinea". In: *Frontiers in Global Women's Health*. URL: https://www.frontiersin.org/articles/10.3389/fgwh.2021.655920/full.

Barman, S. (2012). "Socio-economic and demographic determinants of unmet need for family planning in India". In: *Research on Humanities and Social Sciences* 2.9, pp. 1–12.

Ciganda, D. and N. Todd (2021). "Demographic models of the reproductive process: Past, interlude, and future". In: *Journal of Population Studies*. DOI: 10.1080/00324728.2021.1959943. URL: https://doi.org/10.1080/00324728.2021.1959943.

Devaraj, K., J. Gausman, R. Mishra et al. (2024). "Trends in prevalence of unmet need for family planning in India: patterns of change across 36 States and Union Territories, 1993–2021". In: *Reproductive Health* 21.48. DOI: 10.1186/s12978-024-01781-6. URL: https://doi.org/10.1186/s12978-024-01781-6.

Jena, S.K. et al. (2024). "Demographic and socio-economic correlates of knowledge of the ovulatory cycle among tribal women in India". In: *BMC Public Health* 24.766. DOI: 10.1186/s12889-024-18296-1. URL: https://doi.org/10.1186/s12889-024-18296-1.

K. S. James et al. (2021). *National Family Health Survey (NFHS-5), 2019–21: India*. Mumbai. URL: https://dhsprogram.com/pubs/pdf/FR375/FR375.pdf.

Maricic, M., G. Stojanovic, V. Pazun et al. (2021). "Relationship between socio-demographic characteristics, reproductive health behaviors, and health literacy of women in Serbia". In: *Frontiers in Public Health* 9. DOI: 10.3389/fpubh.2021.629051. URL: https://doi.org/10.3389/fpubh.2021.629051.

Nejhaddadgar, N., A. Mardi and M. Zare (2025). "The impact of new population policies on women's reproductive health in Iran". In: *Frontiers in Reproductive Health*. URL: https://www.frontiersin.org/articles/10.3389/frph.2025.1520027/full.

New, J.R., N. Cahill, J. Stover et al. (2017). "Levels and trends in contraceptive prevalence, unmet need, and demand for family planning in India". In: *The Lancet Global Health* 5.3, e350–e358. DOI: 10.1016/S2214-109X(17)30033-5. URL: https://doi.org/10.1016/S2214-109X(17)30033-5.

Oyinlola, F.F., J.A. Kupoluyi and O.M. Adetutu (2024). "Changes in unmet need for family planning among married women of reproductive age in Nigeria: A multilevel analysis". In: *PLOS ONE*. DOI: 10.1371/journal.pone.0306768. URL: https://doi.org/10.1371/journal.pone.0306768.

Pasa, K., N.M. Singh, N. Bharali et al. (2024). "Reproductive Health Status in India: Comparison of NFHS-IV and NFHS-V Findings". In: *Journal of the Anthropological Survey of India*. DOI: 10.1177/2277436X241282596. URL: https://doi.org/10.1177/2277436X241282596.

Pawar, R.D. and S.P. Solanke (2020). "Unmet needs for family planning and socio-demographic correlates in rural India". In: *International Journal of Community Medicine and Public Health* 7.2, pp. 659–664. DOI: 10.18203/2394-6040.ijcmph20200445. URL: https://doi.org/10.18203/2394-6040.ijcmph20200445.

Prasad, R.V., J. Venkatachalam and Z. Singh (2016). "Unmet needs of family planning among women in rural Tamil Nadu". In: *Journal of Obstetrics and Gynaecology of India* 66.S1, pp. 488–493. DOI: 10.1007/s13224-016-0854-6. URL: https://doi.org/10.1007/s13224-016-0854-6.

Rahaman, M., M.J. Rana et al. (2022). "Spatial heterogeneity and socio-economic correlates of unmet need for spacing contraception in India". In: *Clinical Epidemiology and Global Health* 15, p. 101012. DOI: 10.1016/j.cegh.2022.101012. URL: https://doi.org/10.1016/j.cegh.2022.101012.

Rahaman, M., R. Singh and P. Chouhan (2022). "Levels, patterns and determinants of using reversible contraceptives for limiting family planning in India". In: *BMC Women's Health* 22.124. DOI: 10.1186/s12905-022-01706-0. URL: https://doi.org/10.1186/s12905-022-01706-0.

Sharma, H., S.K. Singh and S. Srivastava (2021). "Spatial heterogeneity and major correlates of unmet need among young married women in India". In: *SAGE Open* 11.2. DOI: 10.1177/21582440211024615. URL: https://doi.org/10.1177/21582440211024615.

Singh, S.K., G.C. Kashyap, H. Sharma et al. (2023). "Changes in discourse on unmet need for family planning among married women in India". In: *Scientific Reports* 13.20464. DOI: 10.1038/s41598-023-47191-9. URL: https://doi.org/10.1038/s41598-023-47191-9.

Zhu, X., Z. Zhu, L. Gu et al. (2022). "Prediction models and associated factors of fertility behaviours of the floating population in China". In: *Frontiers in Public Health*. DOI: 10.3389/fpubh.2022.977103. URL: https://doi.org/10.3389/fpubh.2022.977103.