Prediction of Extra-marital Affairs using Deep Learning Techniques

MSc Research Project
MSc in Data Analytics

Rohan Mohite
Student ID: x23118865

School of Computing
National College of Ireland

Supervisor: Furqan Rustam

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Rohan Mohite
Student ID:	x23118865
Programme:	MSc in Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Furqan Rustam
Submission Due Date:	12/20/2024
Project Title:	Prediction of Extra-marital Affairs using Deep Learning Techniques
Word Count:	xxx
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:	Rohan Sunil Mohite
Date:	16th September 2024

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):	

Prediction of Extra-marital Affairs using Deep Learning Techniques

Rohan Mohite
x23118865

Abstract

Extramarital affairs usually has a huge effect on the involved parties and hence predicting the occurrences would be helpful in counseling. Extramarital affairs in a marital relationship poses severe emotion and social effects, leading to relational deterioration and high cases of divorce. It is therefore important to anticipate such incidences in order to design practical counseling and intercessions. This research tries to fill this gap by using deep learning approach to predict extra marital affair using demographic background, personality traits and socio-economic status. Two predictive approaches were employed: usually distinguished into binary classification and multiclass classification. Some of the pre-processing steps taken were feature scaling, detection of outlier and feature encoding in particular the categorical variables were encoded using one-hot encoding. Among the models used, there are LSTM networks, CNN, Random Forest and SVM. Evaluation paramaters were accuracy, precision, recall, as well as F1 score. The LSTM model performed for binary classification passing with an accuracy of 91.8%, while the CNN model acheived 80.6% for multiclass classification on the same data. These results reveal the potential of deep learning methods for the effective prediction of extramarital affairs and useful knowledge to be applied to relationship counseling and advancement of preventive measures to optimise relationship outcomes.

1 Introduction

Extramarital affairs, sometimes referred to as marital infidelity, are an interesting topic of research in psychology, sociology, and relationship counseling. Despite the sacredness of marriage vows, adultery worldwide still affects a large percentage of couples. Many cultural factors contribute to this. It is important to understand the determinants of extramarital affairs for several reasons Selterman et al. (2021). First of all, it contributes to the development of prevention strategies, which enhance healthy relationships and reduce the frequency of marital conflict and divorce. Secondly, it provides an understanding of the underlying issues as it can lead to adultery in married life without dissatisfaction, emotional distance, or unfulfilled psychological desires Fair (1978). Solving these problems allows couples to work towards a happier and more healthy relationship. There are many different factors that affect the likelihood of an extramarital affair. Nath (2011) research has found many important factors, including personality traits, marital satisfaction, historical trends, age, sex, other socioeconomic factors and other demographic factors have been discovered to be massive predictors by research. For example, it has been shown that younger people and those who have had more sexual partners in the past are more likely to be unfaithful Le et al. (2024). Furthermore, low marital satisfaction and narcissistic and impulsive personality traits are strong predictors of extramarital encounters Namazi and Kermani (2023).

An essential element in predicting infidelity is demographics. Khalili et al. (2021) research indicates that men are regularly more susceptible than women to have extramarital affairs, yet there appears to be a last gender distinction in recent years. It also has an impact on probability of infidelity; people with higher income or partners with financial problems are likely to cheat Dew et al. (2022). Moreover, the results underscore the role of culture and social context on the outcomes of extramarital affairs pointing to the fact that understanding the context is a very relevant factor to consider when studying this process. Likelihood of extramarital affairs is best understood because personality traits are vital in its prediction. According to Rosaline and Kasaraneni (2024) IRIKA and OLAITAN (2018) study it was established that the chances of adulterous behavior was determined by high extraversion and low introversion, in addition, high factor of openness in comparison to low. A person's attitude toward opportunity-taking and new experiences may the influence of these attitudes. They are predicted primarily by relationship satisfaction. Unhappy people in marriage often experience conflict, emotional withdrawal, and sexual dissatisfaction, leading them to seek emotional or physical intimacy after their union. Unresolved problems in marriage can create an environment people are likely to break up in an attempt to satisfy unfulfilled desires elsewhere and can be a barrier against extramarital affairs. The claim that "past behavior is strongly predictive of future behavior" also applies to extramarital affairs. Individuals who have cheated on their spouse in the past show a tendency to do so again in the future. This relapse can be caused by many factors, including repetitive behaviors, failure to address a deeper underlying problem, or lack of remorse.

1.1 Research Motivation

The reason for this research stems from the understanding of how infidelity affects people and the culture. Other social problems which emanate from marital infidelity include tendencies towards the high incidences of divorce, broken families and emotional instability (Fekoor et al. (2024) Nafisa and Ratnasari (2022). This problem should be brought to the attention of everyone so that societies and our relationships can be healthier. Therefore, this paper demonstrates that the use of deep learning techniques provides a fresh angle to the analysis and modeling of infidelity. In contrast to the conventional data analysis methods, deep learning models embrace a huge number of data and develop hidden patterns, that may not be immediately apparent Khalili et al. (2021). This can also imply a change of the effectiveness of relationship counseling by improving both the prognosis capabilities and then, the intervention measures. This study aims at identifying factors that are highly related to extramarital affairs taking into consideration demographic, psychological and socio-economic factors. The idea is to employ these findings towards designing prevention oriented intervention strategies that can be employed to ensure that early warning signs of conflict that may precipitate infidelity in the relationship are dealt effectively.

1.2 Research Objective

The primary objectives of this research are as follows:

- **To identify and analyze key predictors of extramarital affairs**: The purpose of this study is to explore how demographic, personality and socio-economic factors are related to infidelity. Following the progress of the existing literature, this study aims at establishing the relative important factors that predict infidelityLe et al. (2024) Dew et al. (2022).

- **To develop and evaluate predictive models using deep learning techniques**: The studies will compare and contrast different machine learning and deep learning models such as CNNs and LSTMs that shall be used in detecting infidelity with the highest accuracy Rosaline and Kasaraneni (2024). Special attention will be given to models that can handle complex and temporal data patterns.

- **To contribute to the development of intervention strategies**: This knowledge will help in the assessment of personalized preventive strategies to help in reducing the incidences of extramarital affairs Fair (1978). By selecting the most appropriate models and applying them to counseling practices, the goal of this study is to improve satisfaction with the marital relationship and foster stronger connections.

Thus, the findings of this research extend the knowledge of the causes which contribute to individuals engaging in extra marital affairs and provides unique tools for assessing and preventing such conduct. This is a groundbreaking paper with great implications for academic and practical development and growth of relationship therapies based on the application of sophisticated machine learning approaches to relationship counseling.

The structure of the thesis is as follows: Section 2 covers Related Work, Section 3 presents the Methodology, Section 4 details the Design Specification, and Section 5 discusses the Implementation. Evaluation is addressed in Section 6, and the thesis concludes with Section 7, which includes the Conclusion and Future Work.

2 Related Work

The literature on marital infidelity reveals complicated elements starting from economic status and psychological motivation to cultural elements and individual variations. This evaluation sums up all the findings from various research to understand the determinants of marital infidelity and the factors influencing marital satisfaction.Characteristics or circumstances that statistically raise the probability of someone being adulterous. Personal characteristics, interpersonal dynamics, and socioenvironmental factors are a few examples.

2.1 Socioeconomic and Family Conditions

Nazoktabar (2020) tested the effect of family popularity on marital infidelity in Sari, Iran. The study used the Family Factors Scale and the Marital Infidelity Discussion Scale to analyze information from 200 members. The outcomes indicated a widespread relation among income, sociocultural, educational, and religious factors, and infidelity. With economic factors being the most powerful predictor. The study highlights the significance of family situations in emphasizing the risks of infidelity. Similarly, Kundu et al. (2022) conducted a research in Sikkim, India, collected statistics from 240 peoples in rural and urban areas and tested the determinants of extramarital results by use of Poisson and different statistical models. Findings showed that income, financial satisfaction, and marital satisfaction significantly influenced the number of extramarital affairs. Religion and same-sex marriage highlighted how cultural and social variables contribute to a decline in marital faithfulness and adultery.

In expanding on this view, Dall'Agnola and Thibault (2021) discussed a role of information and communication technologies (ICTs) in provoking divorce and extramarital affairs in Kazakhstan. They discovered the fact that internet information use in this or that degree affected the perception of divorce and extramarital affairs issues, these changes was most considerably observed among university-educated young women in Kazakhstan. However, younger and older people tend to maintain traditional marital responsibilities. Age, religion, economic status and education are significant predictors of marital satisfaction in 33 countries, Dobrowolska et al. (2020) in their comprehensive study. This extensive study showed that country-specific characteristics do not have as much of an impact on marital satisfaction as interindividual differences do, indicating that increasing one's personal financial situation may lower the likelihood of infidelity globally.

2.2 Socio-Cultural and Ethical Perspectives

IRIKA and OLAITAN (2018) therefore discussed the socio-cultural and moral implication of extramarital affairs in Nigeria which includes the changes in the family, cultural beliefs in Infidelity, and rate of affairs in the society. Besides the analysis of how culture as well as ethical standards affect attitudes towards infidelity, the paper sheds light on the factors that precipitate extramarital affairs including young marriages, wrong reasons to embrace marriage, failure to fit alterations, and lack of emotional and physical comfort. It proposes to treat the problem with answers and ethical guidelines.Nath (2011) statistically tested the hypothesis that the frequency of extramarital affairs is driven by their numbers. The findings display strong relationships among demographic variables like age, gender, and socioeconomic level, and the prevalence of infidelity.

2.3 Psychological and Emotional Factors

Dehaji et al. (2020) examined the effects of psychological factors and marital satisfaction on the extramarital relationships of married women in Yazd, Iran. Studies using the Symptom Checklist 90 (SCL-90) and the ENRICH marital satisfaction scale revealed significant associations between mental health issues such as anxiety, depression, psychosis, and infidelity. Marital dissatisfaction was also shown to increase the likelihood of extramarital affairs. Implementing mental therapy interventions may reduce infidelity risk and increase marital satisfaction. Accordingly, Hajare (2020) mentioned the dangers adultery poses to physical health, with specific reference to coronary heart diseases in Indian women who have extramarital affairs. The study linked the stress and trauma of infidelity to increased heart disease, highlighting the severe emotional and physical consequences.

Babaei and Zavrei (2021) compared the mental profiles of women and men who have had extramarital affairs and those who have not. Using the Minnesota Multidimensional Personality Questionnaire, significant differences in social inhibition and hypomania were found between those involved and those who were not involved in infidelity. Mtenga et al. (2018) conducted a study in rural Tanzania that examined postmarital status prevalence of relationships and affecting factors among persons having HIV. The study focused on socioeconomic, cultural, and health factors contributing to mistrust in this population.

2.4 Socio-Cultural, Machine Learning and Ethical Perspectives

Rosaline and Kasaraneni (2024) pointed out how the COVID-19 epidemic extended the number of extramarital affairs, with a unique emphasis on troubles associated with work from home. According to the survey, shifting daily pattern, an boom in on-line communications, and a fuzziness of the lines setting apart private and professional lifestyles are all contributing elements to the growth in adultery. Khalili et al. (2021) aimed at identifying the relationships between wives of men in Tehran and their understanding of extramarital affairs, the inferiority of women, manipulation, and religious practices. They had established a positive relationship between perception on extramarital affairs, feeling of locus manipulation and inferiority.

Roman (2020) did a study about online infidelity among the female community of Romania, where the author identified some causes including lack of emotional fulfillment, and the convenience of the internet as some of the leading causes. The focus is laid upon the interdependent nature of personal, relational, and technological conditions, as well as the increased rates of cyber cheating and its consequences for the future types of relationships. Vowels et al. (2022) described the application of machine learning to identify possible predictors of adultery stating the need for AI models' interpretability for the reasons of extramarital affairs. It is as a result of these findings that marital dissatisfaction and personality traits emerges as important correlates of infidelity.

Based on the observation of how people allocate their time between the employment, a spouse and an extramarital companion, Fair (1978) developed an economic understanding of affairs. This method is wise in providing understanding of the reasonable decisions leading to adultery and suggest financial undertakings in order to minimize probability of it.

2.5 Interventions and Solutions

Many works describe unique ways of resolving conflict situations observed in case of infidelity. There are strategies in identifying the underlying issues of the affair, where one can learn to take responsibility of the mistake made by each of the couple and another one is re-establishing and enhancing the mutual respect and understanding between the two partners. Concerning the issue of infidelity, stress is placed on the fact that applying the family therapy and counselling procedures is crucial for repair and prevention of the further enactment. Such interventions cover repair of trust, working on interpersonal relations, and working on mental disorders causing adultery.

IRIKA and OLAITAN (2018) recommended that counseling and ethical education should be used as ways of minimizing the incidences of extramarital interactions. The authors also talked about the role of counseling helping to maintain the marital integrity and working through the dissatisfaction subconsciously present in the marriage. Roman (2020) and her colleagues focued on the significance of paying attention to the relationships and psychological states into account when aiming at reducing the likelihood of cheating. They offered suggestions for strengthening partner accountability and empathy in relationships, highlighting the role therapy plays in preventing future adultery.

Hajare (2020) discussed the information about the high risks for emotional and physical well-being outcomes like a higher risk for cardiovascular diseases, the author pointed the necessity to control the health consequences as well as the emotional and relational ones. This highlights the need for comprehensive measures, including health support. Roman (2020). emphasized the importance of the role of technology and on-line anonymity in facilitating extramarital affairs, suggesting that interventions ought to encompass teaching individuals the risks of on-line infidelity and promoting accountable use of the Internet. Vowels et al. (2022) proposed the use of predictive analytics to identify individuals at risk of adultery and develop early intervention programs. Research has showed that making certain program with respect to particular individual will help to prevent as well as to create an awareness about the risk of adultery and developing a culture in the society that values marital honesty with each other. It is necessary to address underlying psychological and socioeconomic concerns in order to lower the rate of infidelity and promote good partnerships. It is advised that policies be implemented to improve mental health services, financial security, and marital education in order to mitigate the negative impacts of adultery.

The literature review provides a comprehensive understanding of the multifaceted nature of adultery. Important factors affecting this includes modern technology, mental health, marital satisfaction, and financial status. Multiple techniques, together with psychotherapy programming, social assessment, and predictive evaluation, are needed to deal with adultery. More research is needed to study the upward push in infidelity within the digital age.

2.6 Gaps and Limitations

There are some limitations and the lack of certain aspects in the current scholarly research on marital infidelity as identified in the related work. Many studies rely heavily on self-reported data, which can introduce biases and inaccuracies Dehaji et al. (2020) IRIKA and OLAITAN (2018), and the quantitative method itself does not provide an understanding of the personal motives for infidelity Kundu et al. (2022) Fair (1978). Additionally, there is a problem of geographical scope as most of the conducted research is culture-sensitive and has been conducted in certain regions, such as Iran, India, or Nigeria Nazoktabar (2020) IRIKA and OLAITAN (2018) thus restricting the general applicability of the results in other parts of the world. Hence, more comparative research on infidelity in different cultures of the world is required to define if there are cross-cultural similarities and differences, as well as influence of processes like globalization and intercultural marriages. While some contemporary works addressed the contributions of technology to infidelity Rosaline and Kasaraneni (2024) Roman (2020) there are still many unknown long-term effects of digital media on marriages. Also, the application of machine learning models for the identification of the risk of infidelity innovatively Vowels et al. (2022) but it creates the following ethical issues that warrant further discussion. There are many areas in – essence, all areas – where there is a need for more research, for example: The influence of psychological/ emotional factors like the following: mental health disorders and their impacts on the marital relationship and satisfaction Babaei and Zavrei (2021). Counseling and therapy such interventions have not been researched in terms of the impact with different population samples, and more randomized control trials are needed to determine the efficacy of these approaches especially in the long-term. Also, there is the lack of intersectional analysis that aims to look at race, gender, class and sexual factors as they are involved in the experience of infidelity. Filling these gaps with more diverse approaches, cultural paradigms, and psychological and technology based approaches will result in better ways of intervening and a better understanding of marital processes. The Summary of recent literature is provided in table 1.

Authors	Year	Method	Findings
Maryam Fekoor, Alireza Khodami & Fakhrosadat Piltan	2024	Sociological Analysis	Sociological factors, including cultural influences and socio-economic conditions, play a significant role in extramarital relationships among couples.
Tam-Tri Le, Ruining Jin, Minh-Hoang Nguyen & Quan-Hoang Vuong	2024	Survey, Statistical Analysis	Found that moral senses significantly affect young males' attitudes toward extramarital affairs, with an emphasis on generational differences in moral views.
Seyedeh Afsaneh Namazi & Maryam Yavari Kermani	2023	Descriptive-Correlational, Multiple Regression Analysis	Found that sexual assertiveness and spouse selection criteria are significant predictors of the inclination towards extramarital relations in married women.
Haifa Nafisa & Yudiana Ratnasari	2022	Cross-Sectional Study, Multiple Regression Analysis	Marital satisfaction and gratitude significantly predict attitudes towards infidelity among married individuals.
Jeffrey P. Dew, Matthew T. Saxey & Alison Mettmann	2022	Statistical Analysis, Surveys	Identified a relationship between financial deception and extramarital infidelity, with both factors often co-occurring in married relationships.
Ruma Kundu, Kul Bahadur Chettri, Girijasankar Mallik	2022	Poisson Regression, Zero-Inflated Poisson (ZIP), Instrumental Variable Poisson (IV Poisson)	Identified income, financial satisfaction, and marital satisfaction as positive factors for extramarital affairs, with infidelity peaking around 12-13 years of marriage.
Salini Rosaline & Himajyothi Kasaraneni	2022	Systematic Literature Review, Bibliometric Analysis	Found that work-from-home periods increased extramarital affairs due to more online communication and less social interaction.
Laura M. Vowels, Matthew J. Vowels, Kristen P. Mark	2022	Machine Learning (Random Forest, Shapley values)	Identified relationship satisfaction, love, desire, and relationship duration as major predictors of infidelity, suggesting early intervention in relationship problems.
Dylan Selterman, Justin R. Garcia & Irene Tsapelas	2021	Mixed Methods, Survey Analysis	Explored the behavioral, emotional, and sexual outcomes of infidelity, with significant correlations between extradyadic infidelity motives and relationship outcomes.
Mahnaz Babaei & Mahshid Zavrei	2021	MMPI, Statistical Analysis	Found higher levels of hypochondria, social deviation, and hypomania in individuals involved in extramarital affairs compared to those not involved.
Jasmin Dall'Agnola & Hélène Thibault	2021	Mixed-method approach, World Values Survey, focus groups	Concluded that internet use increases approval of extramarital affairs and divorce, with generational differences.
R. Khalini, B. Bahmani, M.S. Khanjani, M. Vahedi	2021	SCL-90, Inferential Statistics (Spearman's Correlation)	Found significant positive relationships between permissive attitudes towards extramarital affairs, feelings of inferiority, and locus of control.
Noraste Semsar-e Dehaji et al.	2020	SCL-90, ENRICH Marital Satisfaction Scale, Statistical Analysis	Found significant links between interpersonal sensitivity, depression, psychosis, and extramarital relations, with marital satisfaction negatively correlated.
Hossein Nazoktabar	2020	SPSS-21, Family Factors Scale, Marital Infidelity Tendency Scale	Found that economic status, social class, culture, education, and religion are significant predictors of marital infidelity.
Diana Roman (Filimon)	2020	Qualitative Research, Content Analysis, MAXQDA	Found that internet infidelity among Romanian women is influenced by personal problems, relational disputes, and sexual concerns.
Sally M. Mtenga et al.	2018	Mixed Methods (Quantitative: Logistic Regression, Qualitative: Semi-structured Interviews, FGDs)	Found a statistical relationship between extramarital affairs and HIV, influenced by cultural and economic factors.

Table 1: Summary of related work on extramarital affairs.

3 Methodology

This chapter provides a comprehensive explanation of the methodology adopted for the study. The primary objective is to predict the occurrence of extramarital affairs using a binary classification approach. The methodology is as follows starting with data preprocessing, handling outliers, defining features and target variables, converting categorical data, splitting the dataset, standardizing the data, and training and evaluating various machine learning models, alongside deep learning algorithms. The proposed methodology is shown in Figure 1.

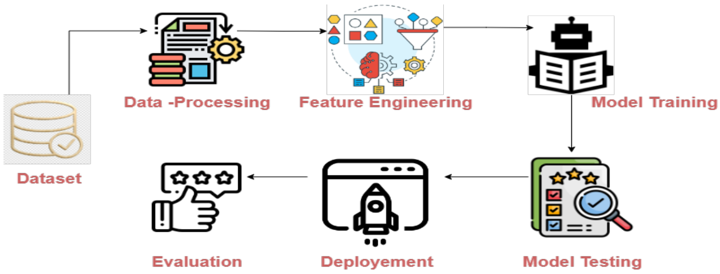


Figure 1: Methodology.

3.1 Data Overview

There are total 13,544 entries organized into 10 columns make up the dataset used by this study, which offers a wealth of information for analysis. The columns representation is showned in Table 2.

Column	Non-Null Count	Dtype
Index	6366 non-null	int64
rate_marriage	6366 non-null	int64
age	6366 non-null	float64
yrs_married	6366 non-null	int64
children	6366 non-null	int64
religious	6366 non-null	int64
educ	6366 non-null	int64
occupation	6366 non-null	int64
occupation_husb	6366 non-null	int64
affairs	6366 non-null	int64

Table 2: DataFrame Summary.

3.2 Data Preprocessing Data Loading and Initial Processing

A Python data processing tool called pandas was used to load the dataset. In addition to the target variable "affairs," which denotes the number of extramarital affairs a person had, the dataset also includes a number of demographic and sociological characteristics. The value variable 'affairs' needs to be converted to binary format. Values greater than 1 were set as 1. This change was necessary to convert the problem into a binary classification task, thus simplifying the predictive modelling process.

	Index	rate_marriage	age	yrs_married	children	religious	educ	occupation	occupation_husb	affairs
count	6366.000	6366.000	6366.000	6366.000	6366.000	6366.000	6366.000	6366.000	6366.000	6366.000
mean	3182.500	4.110	29.029	9.262	1.413	2.426	14.210	3.424	3.850	0.698
std	1837.850	0.961	6.848	7.173	1.481	0.878	2.178	0.942	1.346	2.218
min	0.000	1.000	17.500	1.000	0.000	1.000	9.000	1.000	1.000	0.000
25%	1591.250	4.000	22.000	3.000	0.000	2.000	12.000	3.000	3.000	0.000
50%	3182.500	4.000	27.000	6.000	1.000	2.000	14.000	3.000	4.000	0.000
75%	4773.750	5.000	32.000	17.000	3.000	3.000	16.000	4.000	5.000	0.000
max	6365.000	5.000	42.000	23.000	6.000	4.000	20.000	6.000	6.000	58.000

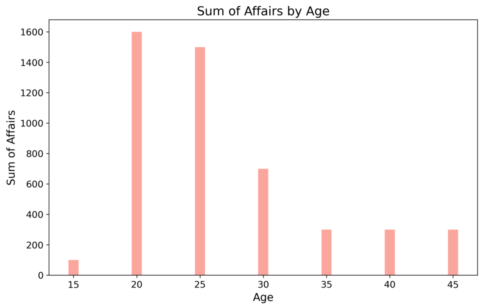
Table 3: Descriptive Statistics.

3.3 Visualization

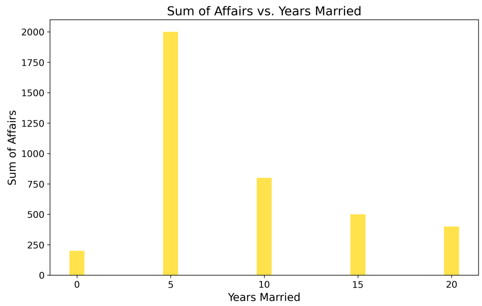
The visualization in this study provides a comprehensive overview of factors affecting extramarital affairs, highlighting key demographic relationships.

3.3.1 Bar Charts

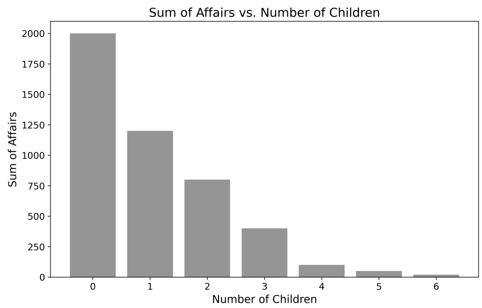
According to the "Affairs vs. Age" graph, people in their 20s and 30s are more likely to have affairs. The "Affairs vs. Years Married" graph shows that affairs rise after age fifteen and fall thereafter, and the "Affairs vs. Number of Children" graph shows the importance of family planning because those without children have the greatest number of affairs whereas It decreases with having more number of children Nath (2011). All this is been represented in Figure 2.



(a) Age vs Affairs.



(b) Years Married vs Affairs.

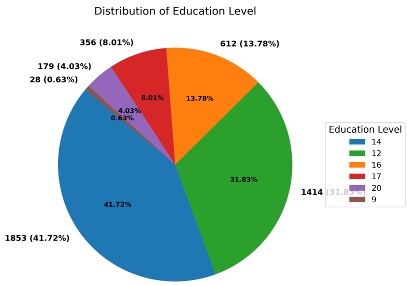


(c) Children vs Affairs.

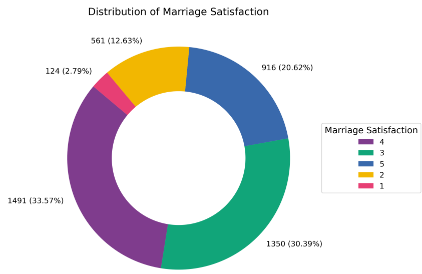
Figure 2: Bar charts of affairs against different factors.

3.3.2 Pie Charts

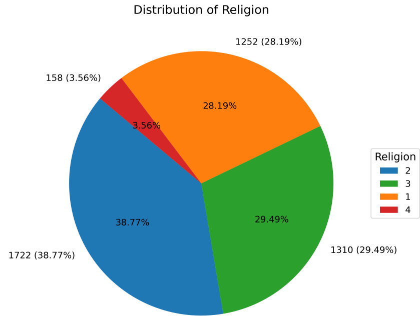
The distribution of extramarital relations among different demographic groups is seen in the pie charts. The majority of respondents have 14 years of education, according to the "Religion vs. Education" chart. The bar graphs show that fewer extramarital encounters are correlated with higher levels of education and marital satisfaction. According to the "Affairs vs. Education Level" graphic, people who have completed 16 years of schooling are the most likely to have affairs, followed by those who have completed 12 and 14 years. Finally, the "Affairs vs. Religion" graphic illustrates the cultural and religious implications by showing that some religious groups have higher incidence of affairs Nath (2011). The visulization can illustrated in Figure 3.



(a) Affairs vs Education.



(b) Affairs vs Marriage Level.



(c) Affairs vs Religion.

Figure 3: Pie charts.

3.4 Handling Outliers

Machine learning models can be severely corrupted by outliers because they distort the training process and produce unreliable predictions. The Z-score approach was used in this. Z-scores had been calculated for every of the statistical data inside the dataset. If the absolute Z-score of a data point us more than three, an outlier was detected and removed from the dataset. This step ensured the data used to train the models is clean and free from errors Rosaline and Kasaraneni (2024).

3.5 Categorical Conversion

The target variables were converted to categorical format using the to_categorical function of the TensorFlow Keras utilities. This flexibility was important for training machine learning models that require categorical labels, such as neural networks.

3.6 Data Splitting

The train_test split function from Scikit-learn was used to divide the dataset right into a training (70%) and testing (30%) set. To guarantee that the results outcomes will be repeated, a random state was used. A common practise in machine learning implementations to assess model performance further Jahin et al. (2024).

3.7 Standardization

Feature scaling was executed using StandardScaler from Scikit-learn. Standardization of functions consists of extracting the mean and scaling to unit variance. This step ensured that all factors contributed equally to model's performance and helped the algorithms used to train the models converge faster.

3.8 Data Reshaping

The LSTM and CNN deep learning models required restructuring the data to fit the favored input format. When a third dimension was introduced, the data took the form (number of observations, number of items, 1). This step ensured the data format matched the input specifications of these models so that pattern and dependencies can be identified.

3.9 Model Training and Evaluation

Several machine learning and deep learning models were trained and evaluated using a variety of measures, including accuracy, precision, and recall, in order to assess how well they performed in predicting the number of extramarital affairs.

- **Logistic regression:** The first model for this research was chosen to be Logistic Regression. Based on a logistic function, this is a linear model that represents the likelihood of a binary end result. For binary class problems, logistic regression is a good as well as efficient technique that gives a baseline towards which complex models can be in comparison Nath (2011) Mtenga et al. (2018).

- **Random Forest:** To examine its prediction accuracy for the binary goal variable, a Random Forest classifier was deployed respectively. The ensemble machine learning approach of Random Forest creates more correct and dependable predictions through combining mutiple decision trees into one. It is an effective tool for classification problems since it can handle large datasets with increased complexity and is resistant to overfitting Vowels et al. (2022).

- **K-Nearest Neighbors:** The KNN classifier machine learning algorirhm was deployed and analyzed to understand the overall performance of the model. The KNN analysed and sends a data point into its k-nearest neighbors in the feature space based on majority. In addition, KNN model is easy to apply and interpret, making it ideal for initial model research Jahin et al. (2024).

- **Support Vector Machines:** The SVM model was trained to classify binary target variables by its ability to efficiently handle higher dimensional spaces. SVM works by locating a hyperplane that efficaciously separates the classes inside the feature space. It works specially well when the data isn't always linearly separable, because it makes use of kernel operations to transform the records right into a higher dimensional space Jahin et al. (2024).

- **Gradient Boosting:** Gradient Boosting is an ensemble learning technique that builds weak learners sequentially to enhance overall performance iteratively. Each new learner is skilled to correct the errors made by means of the preceding learner. This approach is robust and flexible, regularly resulting in high predictive accuracy by means of combining the strengths of a multiple weak models Jahin et al. (2024).

- **Convolutional neural networks:** A CNN was also used to take advantage of its ability to identify local patterns in the data. CNNs do very well on tasks involving picture data, where the spatial connection of objects is crucial. Local pattern recognition has, meanwhile, also been applied to time collecting and sequences Solanki and Solanki (2023).

- **LSTM (long-term and short-term memory) networks:** LSTM network was chosen as the deep learning method for this study. Recurrent neural networks (RNNs) such as LSTM can detect long sequences of dependencies. Time series data and production sequences benefit greatly. The LSTM model was designed to address the flow loss problem common to RNNs and to exploit the ability to store information in long sequences to capture the time dependence and patterns of data Solanki and Solanki (2023).

3.10 Evaluation Metrics

Scikit-learn was used to evaluate the overall performance of models by the use of accuracy scores and classification report metrics.The classification report shows the precision, recall, and F1 scores for each algorithm. Comprehensive assessment of model's performance is depicted by these metrics.

Accuracy: The ratio of correctly predicted cases to all cases.

Precision: The ratio of true positive predictions to total positive predictions.

Recall: The ratio of true positive predictions to the actual positive prediction.

F1-Score: The harmonic mean of precision and recall, providing a balanced metric.

Based on these metrics, models were compared to determine which models performed best in predicting the 'affair' of the binary target variables. A thorough approach to data processing, machine learning model training, variable identification, and performance testing is ensured by this systematic methodology.

4 Design Specification

In developing the prediction models of this study, machine learning and deep learning methods are carefully selected in order to optimize the performance. Selected models include traditional machine learning algorithms such as logistic regression, random forests, K-nearest neighbors, support vector machines, and gradient boosting, as well as advanced deep learning models, especially convolutional neural networks and short-term memory networks. Figure 4 shows the design Specification flow.

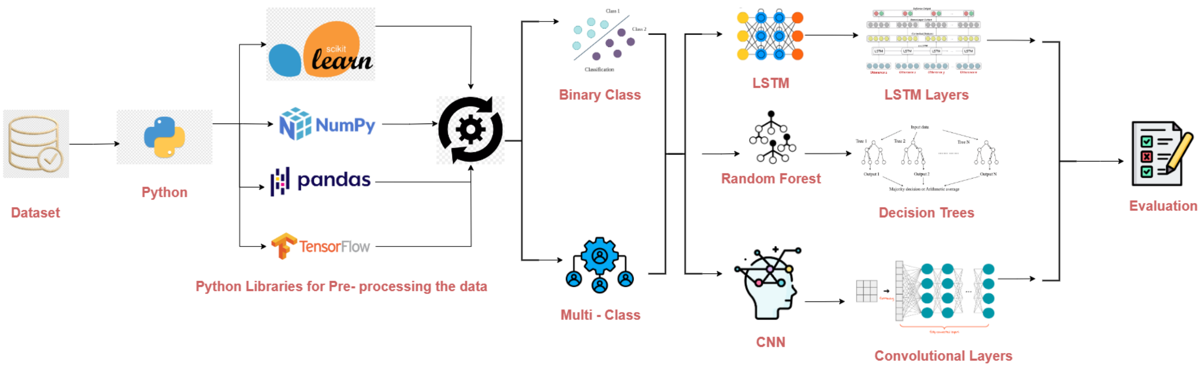


Figure 4: Design Specification Flow Diagram.

4.1 Model with Binary Class

The implementation of the model with consideration of about two classes 0 and 1 present is predicated on numerous key strategies, architectural choices, and frameworks. Methods include data preprocessing, converting the values into binary transformation, feature selection, and hyperparameter tuning. Data preprocessing includes handling missing values using statistical methods such as mean, median, or mode imputation, and normalizing feature values using methods such as the Min-Max scaling or Z-score normalization feature selection ensures the selection of features based on their correlation with target variables. Hyperparameter tuning which makes key parameters such as number of studies, number of estimators, batch size, etc. possible by using cross-validation to model the performance is improved.

The architecture of the model includes defining the structure of the machine learning model. The design uses a specialized machine learning algorithm, such as Random Forest, Support Vector Machine, or Neural Network, with output layers corresponding to class numbers, hidden layers suitable for searching for underlying patterns, and input layers corresponding to the feature number. The model is implemented using machine learning packages based on Python. Scikit-Learn uses traditional machine learning methods like Random Forest and SVM, while TensorFlow/Keras is used to build and train flexible neural network-based models.

The model implementation prerequisites include having enough processing capacity to manage model training ideally with a multi-core CPU and GPU support if neural networks are being used. Additionally, a Python environment with necessary libraries installed (e.g., Scikit-Learn, TensorFlow, Pandas, NumPy) is required, along with a well-structured dataset pre-split into training and testing sets. The Random Forest algorithm, for instance, operates by taking the preprocessed data as input, building an ensemble of decision trees during training, and outputting the class with the majority vote from all trees in the forest for classification.

4.2 Model with Multi Class

With the exception of small changes to account for two specific classes, the best implementation considering the model that takes into account all of the classes that are present is comparable. Methods such as feature selection, data preprocessing, and hyperparameter tuning will remain consistent with the two class method respectively. In data preparation, missing values must be treated accurately in calculations, feature values must be normalized regularly, and features must be selected to ensure comparison. Cross-validation is also used role for hyperparameter adjustment to improve the performance of the model.

The structure of the model is similar to that of the model considering all the classes present, where the main difference is the output layer, namely which corresponds to the reduced number of classes due to the exclusion of two specific classes. The same machine learning algorithms are used to ensure fair comparisons between models. To ensure the consistency of the development process, the model is implemented using the same Python-based machine learning library.

The requirement for the use of the model with two distint class are the same as the model considering all classes present, which include a Python environment with enough computing energy that requires library, training data for the model with two distinct classes, consists of adjusted output levels similar to decreased number of classes. they are always used to ensure comparable performance.

5 Implementation

Developing and testing deep learning and multi-machine learning models to expect the probability of extramarital events is a part of the implementation of the proposed solution. To ensure success with regard to both the structure of the code, modularity, and simplicity, two special approaches binary-class-based and multi-class-based approaches were used. The following table 4 provides a summary of each model's hyperparameter settings.

Model	**Hyperparameters**	**Tuning Range**
Random Forest (RF)	n_estimators = 200 random_state = 23118865	n_estimators = (100 to 500) max_depth = (5 to 20)
Logistic Regression (LR)	random_state = 23118865	solver = {'liblinear', 'lbfgs', 'sag'}
Support Vector Machine (SVM)	kernel = 'linear' C = 1.0, random_state = 23118865	kernel = {'linear', 'rbf'} C = (0.1 to 10.0)
K-Nearest Neighbors (KNN)	n_neighbors = 5, metric = 'minkowski'	n_neighbors = (1 to 10)
Gradient Boosting (GB)	n_estimators = 200 learning_rate = 0.1 max_depth = 3, random_state = 23118865	n_estimators = (100 to 500) learning_rate = (0.01 to 0.3) max_depth = (3 to 10)
LSTM	units = 128 dropout_rate = 0.5 optimizer = 'adam', epochs = 50, batch_size = 32	units = (64 to 256) dropout_rate = (0.2 to 0.5) epochs = (30 to 100)
CNN	filters = 64 kernel_size = 2 dropout_rate = 0.5 optimizer = 'adam', epochs = 50, batch_size = 32	filters = (32 to 128) kernel_size = (2 to 5) dropout_rate = (0.2 to 0.5) epochs = (30 to 100)

Table 4: Hyperparameter Settings for Each Model.

- **Logistic regression:** This model served as a baseline due to its simplicity and difficult interpretation. It used L2 regularization which is form of regularization that has regularization strength of 0.01. Due to its capacity to work with probability interpretations that are binary, logistic regression is most applicable for problems that deal with extramarital attachment probabilities. The model was assessed by accuracy and the model's capacity to classify the data.

- **Random Forest:** An ensemble method with incorporation of 200 decision trees with a specific set random state allowed more generalization, less overfitting and most importantly, the capturing of complex feature interrelationships resulting in improved predictive capability as compared to a single decision tree.

- **K-Nearest Neighbors:** It was considered to be used for classification problems mainly because of the relative simplicity and effectiveness. Therefore, implementing the KNN model with the Minkowski distance metric and a power parameter of 2 (Euclidean distance), the KNN model was set up with five neighbors. In order to predict the results KNN finds the k-neighbors of the target data points and uses the majority classes available among those data points.

- **Support Vector Machines:** In order to handle high-dimensional data, this algorithm was added. The SVM employed a Radial Basis Function (RBF) kernel with a'scale' gamma value of gamma and a regularization parameter (C) of 1.0. SVMs are appropriate for this prediction problem since they work well when there are more dimensions than samples.

- **Gradient Boosting:** An ensemble method that builds models sequentially, each new model correcting mistakes made by using previous ones. The Gradient Boosting model included 200 estimators and a specific random state for consistency. This model is effective for improving accuracy through iterative boosting and combining weak learners to form a robust predictive model.

- **Convolutional neural networks:** The architecture of CNNs was created with the purpose of extracting features from inputs of structured data. The input layer was modified to correspond to a 2D convolutional layer, then ReLU activation function and convolutional layer with 32 filters and 3x3 kernel size were introduced. A MaxPoling layer with 2x2 pooling size was introduced to reduce dimensionality. The features are further extracted and condensed using MaxPooling layer and another convolutional layer with 64 filters and 3x3 kernel size. Before reaching the output layer of the network, which contained a single neuron with a sigmoid activation function, it was flattened and densely layered with 64 layers by ReLU activation The hyperparameters had a learning rate of 0.001, a batch size of 32 and 50 epochs, entropy and loss functions were used respectively.

- **LSTM (long-term and short-term memory) networks:** These networks have been used to locate temporal patterns in sequential data. The architecture consisted of an input layer that reduced to fit with the LSTM inputs parameter, an LSTM layer, and 128 units of ReLU activation.To determine the output of the 2 seeds, a neuron with sigmoid activation was added inside the dense output layer. The graded cross-entropy loss feature and the Adam optimizer were used to train the LSTM model during the course of 50 epochs, with a learning rate of zero.001 and a batch size of 32.

5.1 Tools and Technologies

The models was developed using Python, leveraging its rich ecosystem of libraries for machine learning and data analysis:

- **Python:** The primary programming language used for developing the models due to its simplicity and extensive support for machine learning and data science libraries.

- **Scikit-learn:** This library is majorly applied in the training of several machine learning algorithms like Gradient Boosting, KNN, SVM, Random Forests, and Logistic Regression. Moreover it has the data processing, selection and model analysis functions.

- **TensorFlow:** Utilized to create and train Deep learning models such as CNN and LSTM. Due to its strong backing for the utilization of the GPU, it can process complex graphics and large data sets.

- **Keras:** Neural interfaces can be created and trained using a high level API, which is built on top of TensorFlow.. It facilitates the creation of complex neural networks and pattern training methods.
- **Pandas:** used for data manipulation and preprocessing. It facilitates data cleansing, transformation, and data analytics.
- **NumPy:** Provides support for efficient mathematical calculations and array manipulation. It is important for normalization, encoding and other data preprocessing steps.

These tools and technologies were chosen to ensure efficient development, training, and evaluation of the predictive models, leveraging the strengths of each library to handle various aspects of the machine learning pipeline.

6 Evaluation

The purpose of this section is to provide a comprehensive analysis of the results and key findings of the study, and the implications of these findings from an academic and professional perspective. The most relevant results supporting the research question and objectives are presented, with a strong in-depth analysis.

6.1 Case Study 1: Model with Binary Class

The Multi class-based implementation promoted code modularity and reusability by enclosing the complete machine learning pipeline inside the class. Data preprocessing was handled by the DataPreprocessor class, which handled missing values, normalization of numeric features, and encoding of categorical variables. Each machine learning model was implemented as a separate class, namely LogisticRegressionModel, RandomForestModel, KNNModel, SVMModel, and GradientBoostingModel. These classes include methods for initializing the model, training it on the dataset, making predictions, and testing performance. Deep learning models, especially Convolutional Neural Networks (CNN) and Long-Short-Term Memory Networks (LSTM), were also implemented as classes. These classes included techniques for building network designs, compiling models, training data, and testing performance. Table 5 show the classification report of all the techniques aplied and their respective evaluation.

The LSTMModel and RandomForestModel class, which gives an organisable and reusable structure, used for creation of the LSTM network and trees. 91.8% accuracy, precision of 0.87, recall of 0.94, and F1-score of 0.90 have been obtained with this technique, and similarly for RandomForestModel accuracy of 91.8%, precision of 0.87, recall of 94 and F1-score of 0.90 have been obtained respectively. Table 6 and 7 shows the classification reports of best performing models. Figure 5 shows the confusion matrix and Table 8 shows the cross validation results for binary class.

Class	Precision	Recall	F1-Score	Support
0	0.99	0.90	0.94	1396
1	0.75	0.98	0.85	426
Accuracy	0.92			
Macro Avg	0.87	0.94	0.90	1822
Weighted Avg	0.94	0.92	0.92	1822

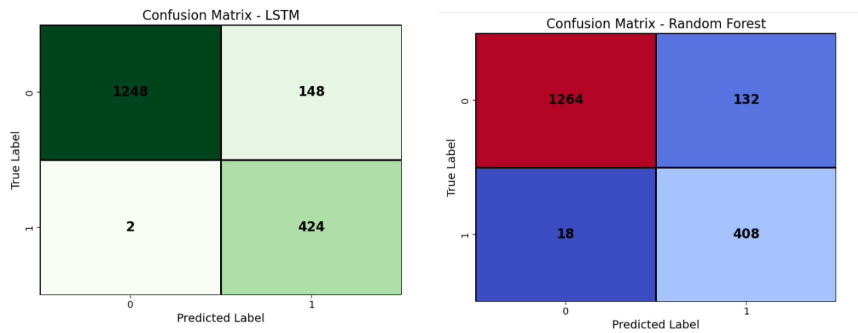
Table 6: LSTM Classification Report.

Class	Precision	Recall	F1-Score	Support
0	0.99	0.90	0.94	1396
1	0.75	0.97	0.85	426
Accuracy	0.92			
Macro Avg	0.87	0.94	0.90	1822
Weighted Avg	0.93	0.92	0.92	1822

Table 7: Random Forest Report.

Algorithm	Accuracy	Class	Precision	Recall	F1-Score
Logistic Regression	89.5%	Class 0	0.94	0.92	0.93
		Class 1	0.76	0.80	0.78
Random Forest	91.8%	Class 0	0.99	0.90	0.94
		Class 1	0.75	0.97	0.85
SVM	90.5%	Class 0	0.96	0.91	0.94
		Class 1	0.75	0.88	0.81
KNN	87.0%	Class 0	0.93	0.90	0.91
		Class 1	0.70	0.79	0.74
Gradient Boosting	91.6%	Class 0	0.99	0.90	0.94
		Class 1	0.75	0.96	0.84
CNN	91.7%	Class 0	0.99	0.89	0.94
		Class 1	0.75	0.97	0.85
LSTM	91.8%	Class 0	0.99	0.90	0.94
		Class 1	0.75	0.98	0.85

Table 5: Comparison of models for Binary Class Implementation.



(a) LSTM Confusion Matrix. (b) RF Confusion Matrix.

Figure 5: Binary Class Confusion Matrix.

Model (Time)	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	Mean Accuracy ($\bar{x}$)	Variance (s^2)
RF (0.59s)	92.20%	92.20%	92.15%	92.05%	92.01%	92.10%	92.26%	91.90%	92.23%	92.12%	1.22E-06
Log R (0.01s)	89.35%	89.36%	89.38%	89.48%	89.31%	89.35%	89.30%	89.38%	89.38%	89.36%	2.38E-07
KNN (0.23s)	86.66%	86.94%	86.99%	87.44%	87.22%	87.49%	87.35%	87.37%	87.49%	87.22%	7.43E-06
SVM (0.16s)	90.75%	90.88%	90.78%	90.78%	90.83%	90.83%	90.86%	90.85%	90.88%	90.82%	1.98E-07
GB (0.93s)	91.65%	91.95%	92.13%	92.28%	92.29%	92.18%	92.36%	92.26%	92.18%	92.14%	4.26E-06
LSTM (2293s)	85.99%	87.42%	85.94%	86.76%	87.88%	86.17%	85.67%	87.52%	87.86%	86.80%	6.97E-05
CNN (3753s)	90.02%	90.01%	90.84%	91.12%	90.84%	90.88%	91.04%	90.96%	90.53%	90.69%	1.57E-05

Table 8: Performance of Different Models across k Values for Binary Class Implementation.

6.2 Case Study 2: Model with Multi Class

In contrast, the implementation of Multi-class based followed a hierarchical approach, where tasks were defined for each step of the machine-learning algorithm With discrete tasks each performed data preprocessing to handle missing values, processing, for statistical accuracy, encode categorical variables and divided the data in training and testing sets. The train_logistic_regression, train_random_forest, train_knn, train_svm, train_gradient_boosting and different features processed the training samples and made predictions. Deep learning models consisting of CNN and LSTM also used netowork structure, model clustering and data training. Analytical metrics including accuracy, precision, take into account and F1 score had been calculated using the dedicated functions. Outputs from the implementation process included preprocessed data, trained models, and analytical proposals. Preprocessed and transformed data are prepared, ready for model training and analysis. The trained models include logistic regression, random forest, KNN, SVM, gradient boosting, CNN, LSTM models Evaluation criteria such as accuracy score, precision, recall, F1 score for evaluation a it goes into detail in the performance of each sample. The evaluation is represented in Table 9.

Algorithms	**Accuracy**	**Precision**	**Recall**	**F1-Score**
LR	79.6%	0.23	0.26	0.24
RF	80.4%	0.32	0.32	0.31
SVM	80.2%	0.18	0.23	0.20
KNN	78.2%	0.26	0.22	0.23
GB	79.1%	0.28	0.28	0.28
CNN	80.6%	0.27	0.29	0.26
LSTM	80.2%	0.20	0.26	0.21

Table 9: Comparison of Models for Multi-Class-Based Implementation.

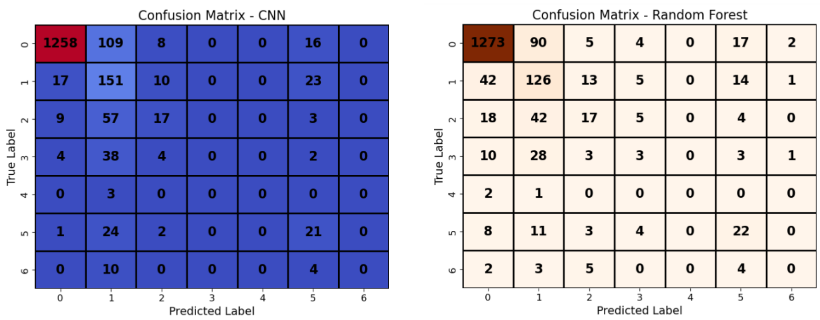
The train_cnn function-controlled Multi-based implementation yielded 80.6% accuracy, 0.27 precision, 0.29 recall, and 0.26 F1-score. Similarly the Random Forest algorithm acheived accuracy of 80.4%, precision and recall of 0.32 and F1-score of 0.31 respectively. Table 10 and 11 shows the detail summary. Figure 6 shows the confusion matrix and Table 12 shows the cross validation results for Multi class Implementation.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.99	0.90	0.94	1391
1	0.38	0.89	0.54	201
2	0.20	0.08	0.12	86
3	0.00	0.00	0.00	48
4	0.00	0.00	0.00	3
5	0.35	0.19	0.24	48
7	0.00	0.00	0.00	14
Accuracy		0.81		
Macro Avg	0.27	0.29	0.26	1791
Weighted Avg	0.83	0.81	0.80	1791

Table 10: CNN Report.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.94	0.92	0.93	1391
1	0.42	0.63	0.50	201
2	0.37	0.20	0.26	86
3	0.14	0.06	0.09	48
4	0.00	0.00	0.00	3
5	0.34	0.36	0.29	48
7	0.00	0.00	0.00	14
Accuracy		0.80		
Macro Avg	0.32	0.32	0.31	1791
Weighted Avg	0.81	0.80	0.80	1791

Table 11: Random Forest Report.



(a) CNN Confusion Matrix. (b) RF Confusion Matrix.

Figure 6: Multi Class Confusion Matrix.

Model	K=2	K=3	K=4	K=4	K=6	K=7	K=8	K=9	K=10	**Mean Accuracy** ($\bar{x}$)	**Variance** (s^2)
RF (8.88s)	80.27%	80.37%	80.52%	80.50%	80.42%	80.60%	80.74%	80.52%	80.55%	80.50%	1.85E-06
Log R (1.27s)	80.08%	80.08%	80.32%	80.13%	80.15%	80.22%	80.13%	80.07%	80.08%	80.14%	6.60E-07
KNN (1.00s)	77.91%	78.16%	78.01%	78.56%	78.43%	78.43%	78.44%	78.64%	78.49%	78.34%	6.51E-06
SVM (4.16s)	80.44%	80.49%	80.44%	80.49%	80.40%	80.49%	80.47%	80.50%	80.52%	80.47%	1.40E-07
GB (82.45s)	79.43%	79.58%	79.98%	79.82%	80.10%	80.02%	80.69%	80.34%	80.30%	80.03%	1.52E-05
LSTM (2191.87s)	78.39%	78.24%	79.46%	78.48%	78.56%	79.60%	78.64%	77.86%	79.18%	78.71%	3.39E-05
CNN (1118.20s)	77.99%	79.92%	79.30%	80.23%	79.97%	79.95%	80.20%	80.15%	80.10%	79.76%	5.18E-05

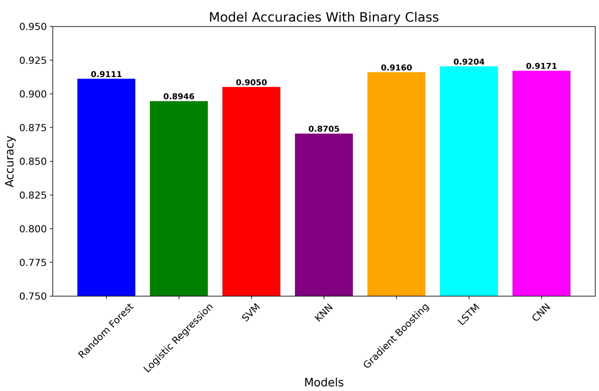
Table 12: Performance of Different Models across K-Folds for Multi Class Implementation.

6.3 Discussion

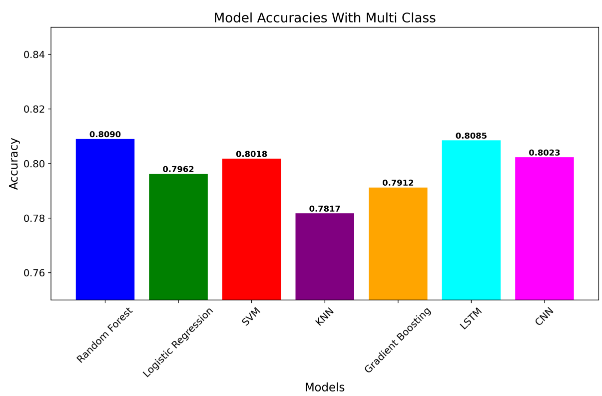
In the binary classification Random forest/LSTM was found as the best models to be implemented. Here, Random Forest outperformed other algorithms in the ability to model complex interaction of the features while providing the robust and accurate estimates. Due to the model's resistance to overfitting, it was ideal for estimating whether the person involved in extramarital affairs. LSTM performed better having an accuracy of 91.8%., it used its memory cell architecture to model sequential dependencies, even when the data did not explicitly contain time-series information. This capability enabled it detect such patterns and to retain relevant information thus contributing to strong results.

For multi-class classification both CNN and RF showed the best result. CNN gave the highest accuracy of 80.6%, employing its convolutional layers for generating complicated hierarchical features. This made it procinct in differentiation between different classes in terms of personality traits, socioeconomic status and other parameters. Random Forest also was stable and utilized an ensemble approach to overcome the issue of class imbalance and classify the instances for multiple classes accurately.

In the process of implementation, preliminary data preparation such as removing data outliers, scaling and encoding of nominal features was performed. Each of the models was tuned for the best hyperparameters according to Table4 including the number of estimators in Random Forest as well as the learning rate in LSTM and CNN. Models were assessed in terms of accuracy, precision, recall and confusion matrix in order to make the study reproducible in future studies and other related applications. Figure 7a and 7b illustrates the accuracy of all the models applied respectively.



(a) Model Accuracy's With Binary Class.



(b) Model Accuracy's With Multi Class.

6.3.1 Critique and Improvement

The cases revealed many areas for improvement. Logistic regression and KNN models can benefit from better handling of class imbalances. Hyperparameter tuning and kernel selection are important for SVM. Random forest and gradient boosting focus on model feature selection and importance. CNN and LSTM models require careful network design and parameter optimization. Hybrid models that incorporate the advantage of numerous techniques have to be investigated in future research. Furthermore, improving data pre-treatment methods such sophisticated feature engineering and handling missing values can also improve the performance. Exploring more sophisticated techniques like ensemble learning and deep learning architectures could further boost predictive accuracy.

7 Conclusion and Future Work

Using deep learning and machine learning algorithms, the study attempted to predict the likelihood of an extramarital affair. Several models were tested in the study, with and without class-based structures, including LSTM, CNN, Gradient Boosting, SVM, Random Forest, Logistic Regression, and SVM respectively.

Key Findings:

- CNN emerged because the best-performing model with an accuracy of 80.6% within the Multi-class-based implementation. Similarly LSTM was best performing model with an accuracy of 91.8% in Binary class based implementation.
- Binary Class-based implementations usually outperformed Multi-class-based ones due to better encapsulation and management of preprocessing and hyperparameters.
- CNN and LSTM models were better in handling complex and sequential records respectively.

Implications: The findings show that Gradient Boosting and other ensemble methods are very effective for this forecasting task. The importance of model structure and hyperparameter tuning was evident in all models. Research highlights the ability of machine learning and deep learning techniques to predict complex social behaviours.

Limitations: Class imbalance and the need for large hyperparameter adjustments are obstacles for the study. Performance of KNN and logistic regression models indicated that data imbalance should be handled carefully. This parameters should be carefully selected in SVM and deep learning models to achieve maximum performance.

Future Work: Hybrid systems taking advantage of multiple algorithms should be an important goal of future research. Testing sophisticated preprocessing techniques such as automated feature engineering and imputation techniques can improve model performance. To further improve prediction accuracy, research on ensemble learning methods, complex deep learning algorithms, such as transformers and attentional processes is recommended, If these models are to be applied to large datasets diversity will also shed light on its simplicity and general applicability.

The capacity for industrial applications is significant, specially in fields along with marital counseling, social research, and focused interventions. Developing consumer friendly interfaces and real-time prediction systems may want to make those models handy to practitioners and researchers, facilitating informed decision-making and personalised support.

References

Babaei, M. and Zavrei, M. (2021). A comparative study of psychological profile of men and women with and without extramarital affairs, *Journal of Research in Psychopathology* **2**(4): 4–9.

Dall'Agnola, J. and Thibault, H. (2021). Online temptations: Divorce and extramarital affairs in kazakhstan, *Religions* **12**(8): 654.

Dehaji, N. S.-e., Chale, S., Dehghani, S., Hedayati, F. S. and Ebrahimi, M. (2020). The role of psychological factors and marital satisfaction in predicting of extramarital relations among married women in yazd, *Journal of Advanced Pharmacy Education and Research* **10**(1-2020): 92–95.

Dew, J. P., Saxey, M. T. and Mettmann, A. (2022). Money lies and extramarital ties: Predicting separate and joint occurrences of financial deception and extramarital infidelity, *Frontiers in Psychology* **13**: 1038169.

Dobrowolska, M., Groyecka-Bernard, A., Sorokowski, P., Randall, A. K., Hilpert, P., Ahmadi, K., Alghraibeh, A. M., Aryeetey, R., Bertoni, A., Bettache, K. et al. (2020). Global perspective on marital satisfaction, *Sustainability* **12**(21): 8817.

Fair, R. C. (1978). A theory of extramarital affairs, *Journal of political economy* **86**(1): 45–61.

Fekoor, M., Khodami, A. and Piltan, F. (2024). Sociological investigation of factors influencing extramarital relationships in couples from bandar abbas, *Journal of Assessment and Research in Applied Counseling (JARAC)* **6**(1): 34–44.

Hajare, R. (2020). Extra-marital affair can lead trauma and connections in higher risk for heart disease in indian women. drug des int prop int j 2 (1)-2018, *DDIPIJ. MS. ID* **126**(1): 2.

IRIKA, J. and OLAITAN, O. (2018). A critical evaluation of extra marital affairs in contemporary nigeria society, *Retrieved from researchgate. net/publication/327339258* .

Jahin, M. A., Naife, S. A., Lima, F. T. J., Mridha, M. F. and Shin, J. (2024). Analyzing male domestic violence through exploratory data analysis and explainable machine learning insights, *arXiv preprint arXiv:2403.15594* .

Khalili, R., Bahmani, B., Khanjani, M. S. and Vahedi, M. (2021). The relationship between attitude towards extramarital affairs with the feeling of inferiority, locus of control, and practice of religious beliefs in wives of addicted men, *Journal of Assessment and Research in Applied Counseling (JARAC)* **3**(1): 28–37.

Kundu, R., Chettri, K. B. and Mallik, G. (2022). Extramarital affairs in sikkim and their causes: A study with emphasis on the difference between the sexes, *International Journal of Management and Humanities (IJMH)* **8**(6): 19–26.

Le, T.-T., Jin, R., Nguyen, M.-H. and Vuong, Q.-H. (2024). Fair but irresponsible: How moral senses affect young males' attitudes toward extramarital affairs, *SAGE Open* **14**(3): 21582440241264190.

Mtenga, S. M., Pfeiffer, C., Tanner, M., Geubbels, E. and Merten, S. (2018). Linking gender, extramarital affairs, and hiv: a mixed methods study on contextual determinants of extramarital affairs in rural tanzania, *AIDS research and therapy* **15**: 1–17.

Nafisa, H. and Ratnasari, Y. (2022). Prediction of attitudes towards infidelity among married individuals based on marital satisfaction and grateful disposition, *3rd Tarumanagara International Conference on the Applications of Social Sciences and Humanities (TICASH 2021)*, Atlantis Press, pp. 1593–1599.

Namazi, S. A. and Kermani, M. Y. (2023). Predicting the inclination to extramarital relations based on sexual assertiveness and spouse selection criteria in married women, *Psychology of Woman Journal* **4**(2): 24–31.

Nath, S. (2011). What makes people infidel? an analysis of the influence of demographics on extramarital affairs, *Undergraduate Economic Review* **8**(1): 5.

Nazoktabar, H. (2020). Predicting marital infidelity based on family conditions of couples referred to social emergency., *Journal of Fundamentals of Mental Health* .

Roman, D. (2020). Justifications of internet infidelity in romanian women with extradyadic affairs: a qualitative study proposal, *New trends in psychology* **2**(1).

Rosaline, S. and Kasaraneni, H. (2024). Surge in extramarital affairs during work from home: What we know so far, *The Family Journal* **32**(1): 95–101.

Selterman, D., Garcia, J. R. and Tsapelas, I. (2021). What do people do, say, and feel when they have affairs? associations between extradyadic infidelity motives with behavioral, emotional, and sexual outcomes, *Journal of Sex & Marital Therapy* **47**(3): 238–252.

Solanki, P. S. and Solanki, Y. K. (2023). Revolutionizing divorce case prediction in india: A machine learning approach to save marriages and enhance decision accuracy, *International Journal of Engineering and Management Research* **13**(2): 232–240.

Vowels, L. M., Vowels, M. J. and Mark, K. P. (2022). Is infidelity predictable? using explainable machine learning to identify the most important predictors of infidelity, *The Journal of Sex Research* **59**(2): 224–237.