Journal of Information Systems Engineering and Management
2025,10(8s)
ISSN:2468-4376
https://www.jisem-journal.com/ **Research Article**

Work – Life Balance among Female Lecturers in Tamil Nadu, India: A Synthetic Data Approach

Logeswari B[1], Sathyapriya J[2*], Sivakami T[3], Panneerselvam K[4], Thamilvanan G[5], Bellarmin Diana S[6], Rajandran KVR[7], Gayathri S[8]

[1]Research Scholar, Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, Tamil Nadu & Assistant Professor, Idhaya College for Women, Kumbakonam, Tamil Nadu, logeswaribaskaran16@gmail.com
[2*, 3,4,7]Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, Tamil Nadu, *sathyapriya@pmu.edu / sivakami@pmu.edu / panneerselvam@pmu.edu / kvrrajan@pmu.edu
[5]Department of Education, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, Tamil Nadu, thamilvanan@pmu.edu
[6,8] Bon Secours College for Women, Thanjavur, Tamil Nadu, dianabellarmin@gmail.com / gayathriselva2010@gmail.com

ARTICLE INFO

Received: 23 Oct 2024
Revised: 14 Dec 2024
Accepted: 29 Dec 2024

ABSTRACT

Work-life balance (WLB) is an important issue of wellbeing, particularly for women in professions such as teaching. The aim of this study is to present innovative ways of using synthetic sources to analyse real-world challenges at the intersection of professional duties and personal responsibilities. This research explores the use of artificially generated information and computational techniques to assess the WLB stability of female teachers in higher education institutions. Using this enhanced dataset, a series of machine learning (ML) classifiers are constructed and evaluated to predict WLB outcomes.
This research makes a significant contribution to the fields of educational administration and data science in the Kumbakonam district of Tamil Nadu and illustrates the value of synthetic data in social science research, providing insights into improving the WLB of women in higher education. It also provides recommendations for legislative and institutional reforms that can be implemented to improve women's WLB.

***Keywords*:** Educational Institute, Female Lecturers, Predictive Data, Work-Life Balance, Machine Learning.

1. Introduction

The equilibrium between professional obligations and personal life is a crucial element of overall well-being, particularly for women in academic roles who frequently manage complex pressures in both domains. Women educators in the field of higher education encounter specific challenges that may significantly impact their capacity to reconcile their professional and personal lives. These issues include the demands of their workload, gender-based prejudice, and familial obligations (Noronha & Aithal 2019a; Noronha & Aithal 2019b; Vasumathi, 2018). Gaining comprehension and enhancing the stability between work and personal life in this particular situation is not only crucial for the educators' welfare, but also for the general efficiency and endurance of academic institutions.
A significant challenge is faced by many individuals, particularly women in demanding sectors such as academia, in balancing their professional responsibilities with their domestic lives. Female educators in higher education frequently assume a multiplicity of responsibilities, encompassing the roles of instructors, scholars, advisors, and caretakers for their families (Shaikh & Chandio, 2024). The necessity to flourish in one's professional pursuits while concurrently managing personal and familial obligations may precipitate stress, exhaustion, and diminished work satisfaction. It is crucial to understand the factors that contribute to achieving work-life balance in this particular group in order to develop effective strategies to enhance their overall well-being and productivity (Pace & Sciotto, 2022; Lekchiri & Eversole 2020).
The concept of work-life balance (WLB) is concerned with achieving a state of equilibrium between the demands of one's professional life and the activities of one's personal life. The capacity of women in higher education to attain a work-life equilibrium is influenced by a complex interplay of factors. These factors include the volume of work, the level of assistance received from the educational institution, familial responsibilities, social expectations, and personal objectives. Despite the significant relevance of this issue, there are numerous obstacles associated with investigating work-life balance (Shabir et al. 2021). The use of

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

conventional data-gathering techniques, such as surveys and interviews, can present certain difficulties, including limited participation, concerns about privacy, and the potential for researcher bias. Furthermore, the sensitive and complex nature of the subject matter may present challenges in obtaining comprehensive and accurate data. The creation of synthetic data represents a novel and efficacious approach to addressing these difficulties. The term 'synthetic data' is used to describe data that has been created with the specific intention of replicating the statistical characteristics of real-world data. Researchers may generate synthetic datasets that are realistic and varied through the use of approaches such as data simulation, generative models, and sophisticated algorithms (Akanji et al. 2020). This strategy addresses the issue of privacy while simultaneously providing a comprehensive dataset for the training and assessment of machine learning models. Advanced approaches, such as Generative Adversarial Networks (GANs), will be employed to generate diverse and realistic synthetic datasets. The datasets will be employed for the training of multiple machine learning classifiers, including Logistic Regression (LR), Decision Trees (DT), Support Vector Machines (SVM), and Neural Network (NN). The efficacy of these classifiers will be evaluated through the measurement of pertinent metrics, including accuracy, precision, recall, and F1-score. The aforementioned metrics will assist in the selection of the most accurate models for predicting work-life balance outcomes (Figure 1).

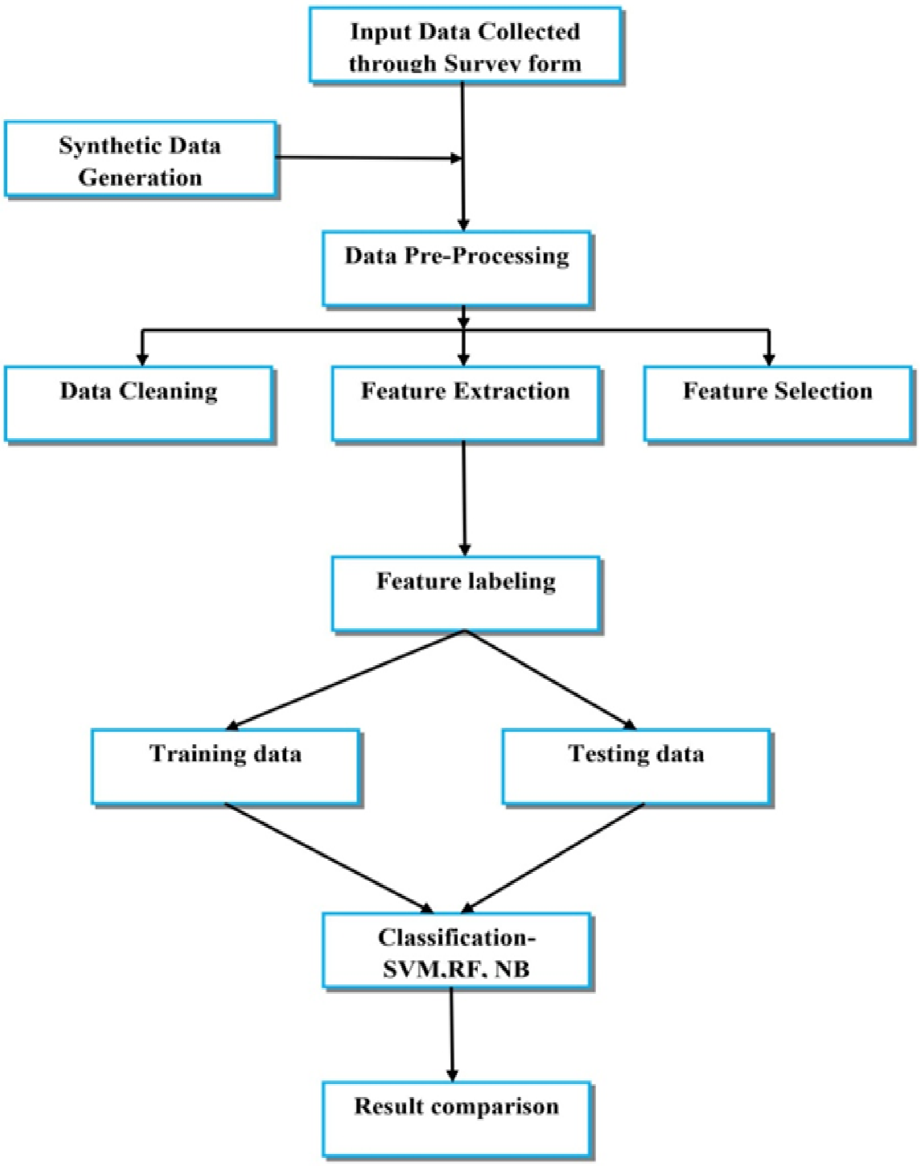

Figure 1: Workflow of proposed study (source: of authors')

The aim of this research is to utilize synthetic data generation and machine learning models to examine and categorize the work-life balance (WLB) of female educators in higher education. The study will commence with a comprehensive literature review to identify the key factors influencing work-life balance within this particular context (Bharadwaj & Shanker 2021). The design of a comprehensive survey instrument will be informed by these variables and will serve to obtain initial data from women educators. The data thus

gathered will serve as the basis for the creation of synthetic datasets that will, in turn, provide an accurate representation of the nuances of work-life balance within this specific group.
The study is expected to yield two specific results. The primary objective is to ascertain the reliability and efficacy of synthetic data as a research tool in the social sciences, particularly in circumstances where actual data is scarce or sensitive. This will be achieved by combining the generation of synthetic data with machine learning (ML). Furthermore, the investigation aims to provide pragmatic and impactful insights into the pivotal aspects that considerably affect the work-life equilibrium of women in advanced education (Pandey & Dhanopia 2019). These perceptions may prompt institutional decision-makers to implement regulations and processes designed to promote the well-being and career advancement of female instructors. Furthermore, the study aims to elucidate the intricate interrelationship between professional obligations and personal responsibilities, and to examine the ways in which organizational culture and leadership can reasonably support women's diverse duties and advancement, through both compassionate accommodation and empowering opportunities (Pandey 2023).

2. Literature review

In their study, Ahmad & Wani (2011) investigate the influence of organizational support on the work-life balance of employed women. There is a dearth of research in this field, both in India as a whole and in Jammu and Kashmir (J&K) in particular. This research employs an empirical analysis to investigate the impact of these three factors on the work-life balance (WLB) of female faculty members at the college and university level in Kashmir. The findings of the research study indicated a significant correlation between all of the criteria and a woman's work-life balance. It is recommended that modern firms implement robust organizational policies to facilitate and advance the career development of women professionals in the workplace.
In a comprehensive assessment conducted by Aafreen et al. (2022), a range of possibilities and problems were revealed. For example, the researchers found that while women may attain high positions in some regions, they are constrained by family responsibilities in other areas. It is therefore essential to fulfil one's familial obligations. In contemporary society, women assume prominent roles within the familial unit, concurrently fulfilling responsibilities within and beyond the domestic sphere. For many career-driven women, achieving equilibrium amidst increasing demands represents a significant challenge. This manuscript focuses on the experiences of working women, examining the relationship between professional fulfilment and personal well-being. By considering a range of perspectives and practical strategies, an optimized framework is proposed which acknowledges the complexity of life while seeking to cultivate peace across its diverse domains.
In their study, Noronha & Aithal (2020) examined the difficulties encountered by female educators in achieving an optimal balance between their professional and domestic responsibilities. The study focused primarily on State Universities in Karnataka. The sample responses are employed to evaluate the hypothesis that a correlation exists between the challenges encountered and the equilibrium between professional and domestic responsibilities. The difficulties encountered have been classified into three domains for examination. The factors under examination are personal characteristics, familial considerations and organizational elements. A total of 422 female instructors from state universities in Karnataka were selected for inclusion in the study. The survey findings indicated that the respondents experienced difficulties to a moderate extent. Furthermore, the findings indicated a correlation between the difficulties and the equilibrium between work and personal life. The findings of this study will contribute to our understanding of the difficulties encountered and their correlation with work-life balance.
In a study conducted by Burke (2001), the relationship between the perception of organizational principles that foster a harmonious work-life balance among female managers and professionals in their workplace, and their work experiences, satisfaction levels in both work and non-work areas, and psychological well-being was examined. The data was obtained from a sample of 251 women via the administration of anonymous questionnaires. Female managers who indicated that their organizations prioritized a work-personal life balance reported higher levels of job and career satisfaction, reduced work-related stress, a lower intention to leave their jobs, improved family satisfaction, fewer psychosomatic symptoms, and enhanced emotional well-being. It was unexpected that there was no correlation between the perception of a company's principles supporting a work-life balance and the number of hours worked, including overtime, or the level of engagement in employment.
Sethi (2014) conducted a comparative study to ascertain the relationship between work-life balance (WLB) and organizational commitment among female employees in public and private sector banks. The assessment of organizational commitment was conducted using a standardized scale created by Mowday, Steers, and Porter (1979). The findings indicated a positive correlation between work-life balance (WLB) and organizational commitment.
Popoola & Fagbola (2020) conducted a study to investigate the influence of work-life balance (WLB), job motivation, and self-esteem on the level of commitment among library workers at federal institutions in southern Nigeria. The researchers employed the comprehensive enumeration technique to include the entire population of 1,138 library professionals employed at federal institutions in southern Nigeria. A standardized questionnaire was administered to 1,138 library professionals, and 1,023 of them completed the

questionnaire and provided replies. A response rate of 90% was achieved. The study revealed that the participants' level of dedication to their company was significantly influenced by a number of factors, including the equilibrium between work and personal life, inspiration in their profession, and their self-worth. The study found that self-esteem is the primary factor influencing organizational commitment, accounting for 34% of its overall effect. It is therefore recommended that library administrators, particularly those involved in education policy, prioritize the enhancement of work-life balance, job satisfaction and self-esteem in order to cultivate a greater degree of dedication among their workforces.

Muafi & Marseno (2021) conducted a quantitative study by administering questionnaires to 118 employees of the Kebumen Branch and Unit of BRI Bank using a purposive sampling method. The data analysis employed structural equation modelling (SEM) with second-order confirmatory factor analysis, utilizing the SmartPLS 3.0 software. The findings indicate that achieving a balance between work and personal life has a considerable and advantageous impact on both work engagement and organizational commitment. Furthermore, emotional intelligence has a considerable and beneficial impact on both job engagement and organizational commitment. Additionally, job engagement has a favorable and substantial impact on organizational commitment. Work engagement serves as a mediator between achieving a balance between work and personal life and the level of dedication and loyalty that an individual has towards their organization. Work engagement also acts as a mediator between emotional intelligence and organizational commitment among employees of the Kebumen Branch and Unit of PT Bank BRI.

In their study, Thi et al. (2021) demonstrated that organizational culture exerts a significant influence on employee satisfaction, which in turn has a substantial impact on their commitment to the firm. The data were derived from a survey of 240 individuals employed in office settings. Structural Equation Modelling is employed to ascertain the impact of a multitude of organizational culture attributes, including a transparent work environment, remuneration and rewards, the delegation of authority, leadership style and corporate principles, on the well-being of employees and their dedication to the company. The findings indicated that the five dimensions of organizational culture exert a positive influence on employees' well-being, while job satisfaction exerts a significant effect on their level of dedication to the firm. Job satisfaction functions as a mediator between business culture and commitment. This study offers a practical understanding of the relationships between organizational culture, job satisfaction among employees, and organizational commitment. The information presented here may be used by managers to develop personnel policies that are appropriate for their companies. The implementation of these regulations will enhance both employee performance and well-being, thereby promoting higher employee retention rates.

In their study, Reddy et al. (2010) investigated several factors that might potentially influence work-family conflict (WFC) and family-work conflict (FWC) among married women employees. The methodology employed in this study is as follows: The sample comprised 90 married women of working age, ranging from 20 to 50 years. The levels of work-family conflict (WFC) and family-work conflict (FWC) experienced by working women were evaluated using the WFC and FWC Scale. The data were subjected to descriptive and inferential statistical analysis. Pearson's correlation coefficient was employed to ascertain the relationship between the various variables.

In their study, Aiswarya & Velmurugan (2022) examine the impact of stress and strain on the work-life balance of financially independent college teachers. The findings of this study have significant implications for the higher education sector. The objective of this study is to elucidate the work-life balance issues faced by self-funded lecturers in Trivandrum. Data for the study was collected through the administration of questionnaires and subsequently analyzed using a range of statistical methods. The findings indicate that self-funded college teachers have a slightly better work-life balance than the overall average.

3. Methodology

The data obtained from the literature survey was subjected to analysis using data mining techniques and statistical procedures. In the initial research contribution, it was employed a range of machine learning classification techniques, including random forest (RF), naive Bayes (NB), and support vector machines (SVM), to analyse and classify the input data. A comparative study was conducted to evaluate the performance of various classifiers, followed by an analysis of the resulting data.

The data was gathered from a sample of employed women residing in the Kumbakonam area using a Google form that was specifically designed for this purpose. The initial dataset comprises 750 discrete responses from a heterogeneous group of individuals, encompassing 25 characteristics pertaining to diverse aspects of their professional and personal experiences. The data underwent a series of standard procedures to ensure consistency and the validity of the responses throughout the cleaning process. In order to enhance the model's specificity, certain attributes have been excluded. In the course of our investigation, we selected 14 characteristics on the basis of their utility. The sophisticated algorithms utilize an employee's historical medical records pertaining to stress-related illnesses in order to make accurate predictions. Furthermore, the "one hot encoding" approach is employed to represent numerous fields that necessitate discrete parameters. The written responses were assigned numerical weights in accordance with their degree of significance. In many cases, the value of 'yes' is assigned a value of 1, 'no' a value of 0, and 'maybe' a value of 0.5. The cells bearing the acronym 'NaN' (meaning 'not a number') were replaced with a value of 0. A label encoder was

employed to transform the category input into numerical values. The training process received 70% of the responses, while the testing process received the remaining 30%.
A total of 750 replies were identified through the application of a machine learning technique, which proved to be an ineffective approach for achieving higher accuracy due to the presence of imbalanced data. A methodology for the generation of synthetic data is presented as a means of achieving data balance. The data was divided into two distinct sets: a training set and a testing set. The training set was allocated 70% of the data for model training, while the remaining 30% was reserved for testing. The utilization of synthetic data facilitates the generation of robust machine learning models, while simultaneously ensuring the confidentiality and anonymity of actual participants.

3.1 Use of Synthetic Data

The use of synthetic data is essential for examining the work-life equilibrium of female educators in higher education for a number of compelling reasons. Firstly, it ensures the confidentiality and security of data by allowing researchers to analyse data that closely resembles real-world information while protecting sensitive personal information (Fonseca & Bacao, 2023). This methodology addresses the issue of limited data availability by creating extensive and varied datasets that may be challenging to obtain through conventional methods due to low response rates and logistical constraints. Furthermore, the utilization of synthetic data assists in the reduction of the intrinsic biases associated with the collection of real-world data, thereby ensuring an equitable and impartial representation of the population. The provision of a substantial quantity of data for machine learning algorithms facilitates the development and testing of sophisticated models, resulting in more precise and reliable predictions. Researchers are able to conduct ethical experimentation, which enables them to investigate hypothetical situations and manipulate certain factors without any ethical considerations. Moreover, the cost-effectiveness and resource efficiency of generating synthetic data make it an attractive option for large-scale research projects (Figueira & Vaz 2022). In conclusion, the use of synthetic data is important for comprehensive and ethical analysis, offering significant advantages in understanding and improving work-life balance among female educators.

3.2.GAN Architecture for Synthetic data

Generative Adversarial Networks (GANs) belong to a category of machine learning frameworks. The main goal of GANs is to produce novel, artificial samples of data that closely match the data used for training. The present paper offers a thorough analysis of the composition and functioning of GANs (Figueira & Vaz, 2022).
The generator ***(G)*** is a neural network that generates synthetic data by using random noise. The main objective is to provide data that is indiscernible from authentic data. The generator takes as input a random noise vector, denoted as z, which is usually sampled from either a Gaussian or uniform distribution. levels: The generator is composed of several levels, such as completely linked layers, batch normalization layers, and activation functions like ReLU or Leaky ReLU. When it comes to generating images, transposed convolutional layers (sometimes referred to as deconvolutional layers) are used to increase the size of the noise vector into a bigger data structure. The generator produces a synthetic data instance that matches the dimensions and format of the actual data.
The discriminator ***(D)*** is a neural network that functions as a binary classifier, differentiating between authentic data and artificial data generated by the generator.
Input: The discriminator receives input that might be either a genuine data instance or a synthetic data instance produced by the generator. The discriminator is composed of many layers, which include convolutional layers (for images), batch normalization layers, and activation functions like Leaky ReLU. The discriminator produces a probability score ranging from 0 to 1, which indicates the likelihood that the supplied data instance is either genuine (near to 1) or synthetic (close to 0).

3.2.1 Synthetic Data Generation

The initialization of the network is the first step in the process. The second step is the adversarial training loop. The first stage of the second step is to train the discriminator (D). A subset of the actual data instances should then be extracted from the training dataset. A set of artificial data instances should be produced using the generator. The loss of the discriminator is calculated, which quantifies its capability to differentiate between genuine and artificial data. The loss is frequently constituted of two components: The cost incurred when accurately categorizing actual data occurrences as genuine (real loss). The loss incurred when categorizing artificial data instances correctly as fraudulent (fake loss). The weights of the discriminator should be optimized using backpropagation in order to reduce the discriminator's loss. The second stage is the training generator (G). The generator should be employed to generate a set of artificial data instances. The generator's loss should be calculated, which serves to quantify the discriminator's performance on the synthetic data. The objective is to optimize the discriminator's loss for artificial data, which indicates that the generator is attempting to deceive the discriminator. The generator's weights should be optimized using backpropagation in order to reduce the generator's loss. The training process continues until the generator generates data that is indistinguishable from genuine data by the discriminator, resulting in the discriminator's accuracy stabilizing at about 50%. This implies that the generator is producing artificial data that closely matches real data (convergence).

The purpose of the generator loss (L_G) is to optimize the discriminator's loss for artificially generated data (1):

$$L_G = -E_{z\sim pz}[\log D(G(z))] \qquad (1)$$

This loss incentivizes the generator to produce data that is indistinguishable from genuine data by the discriminator. The discriminator loss (L_D) is the sum of the loss for actual data and the loss for synthetic data (2):

$$L_D = -E_{x\sim pdata}[\log D(x)] + E_{z\sim pz}[\log(1 - D(G(z)) \qquad (2)$$

In this context, D(x) represents the output of the discriminator when given actual data x, whereas D(G(z)) represents the output of the discriminator when given synthetic data created by the generator using a noise vector z.

3.3 Machine Learning Algorithm

Employing a ML pipeline utilizing synthetic data may proficiently categorize the work-life equilibrium of women in higher education. By doing data preparation, model selection, training, and fine-tuning, one may get accurate and dependable predictions. This methodology facilitates comprehension of the pivotal elements that influence the equilibrium between work and personal life and may provide direction for formulating policies and implementing measures to enhance the welfare of female educators. Data pre-processing is an essential and vital stage in any ML effort. Data preprocessing is the process of converting unprocessed data into a refined and practical format, leading to significant improvements in the efficiency of ML models. This section will provide a comprehensive overview of the data pre-processing procedure. It encompasses handling missing values, encoding categorical variables, normalizing numerical features, and splitting the data into distinct training and testing sets (Appendix – table 1).

Missing values may be addressed by imputing them with suitable statistical measures, such as the mean, median, or mode. In order to streamline the procedure, we will substitute missing values with the mean for numerical features and the mode for categorical attributes (Appendix – table 2). Categorical variables must be transformed into a numerical representation. One-hot encoding might be used for this objective (Appendix – table 3).

It is essential to normalize numerical characteristics in order to ensure that they are comparable on an equivalent scale. Standardization is a widely used method in which each characteristic is adjusted to have an average value of 0 and a standard deviation of 1. This research employed a dataset comprising 1,500 data points, each containing 25 distinct attributes. The dataset comprised a combination of original and generated data. Of the 1,500 data points, 70% (1,050) have been designated for training purposes, while the remaining 30% (450) have been reserved for testing. The objective of this technique is to identify machine learning methods that are appropriate for our specific context. The principal objective is to ascertain the optimal forecasting model for our specific requirements. In order to facilitate analysis, the RF, SVM, and NB classifiers have been considered. Each of the aforementioned classifiers must undergo a process of training and testing.

3.3.1 Classification using Random Forest Classifier

The Random Forest (RF) algorithm is a method of ensemble learning that generates many decision trees during the training process and predicts the class that is most often seen across the different trees (classification). This talk will examine the specific techniques and underlying mathematical concepts used in using Random Forest for the classification of women's WLB in higher education. The RF method provides a comprehensive procedure for identifying WLB among women instructors in higher education using a RF classifier. The essential stages include data preprocessing, training several decision trees using bootstrap samples, combining predictions using majority voting, and assessing the model's performance on a separate test set. RF classifiers provide reliable and precise predictions by harnessing the potential of ensemble learning.

The performance of the RF classifier in WLB classification is shown in the confusion matrix in Figure 2 (and Table 5 – Appendix). The categorization is based on three distinct categories: poor, moderate and good balance. Out of the total incidents, 38 are accurately labeled as poor class, 74 are accurately classified as moderated class, and 46 are accurately classed as good balance of work life.

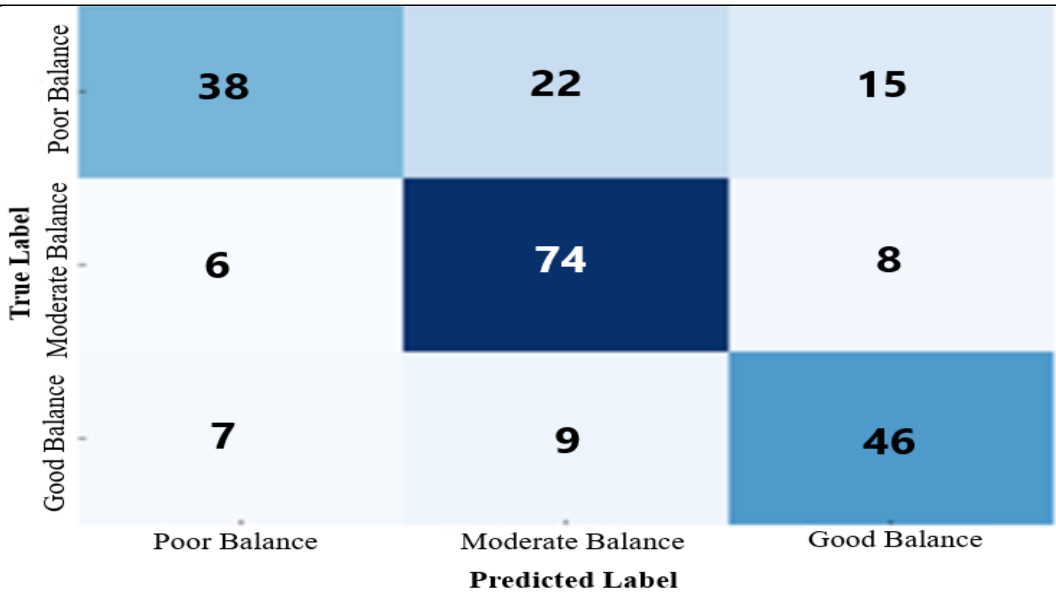

Figure 2: Confusion matrix of RF

3.3.2 Classification using Support Vector Machine Classifier

The technique provided a comprehensive procedure for categorizing the WLB of female professors in higher education using an SVM classifier. The essential stages include data preprocessing, resolving the SVM optimization issue using kernel functions, managing multi-class classification, and assessing the model's performance on a test dataset. SVMs are very effective classifiers, particularly when dealing with data that has a large number of dimensions and requires non-linear decision limits. Each classifier distinguishes between two classes, and the final prediction is made by a majority voting scheme. According to Figure 3 (and Table 6 – Appendix), 90 occurrences were accurately identified as having Poor Balance. 85 cases accurately classed as Moderate Balance, and 100 cases accurately categorized as Good Balance.

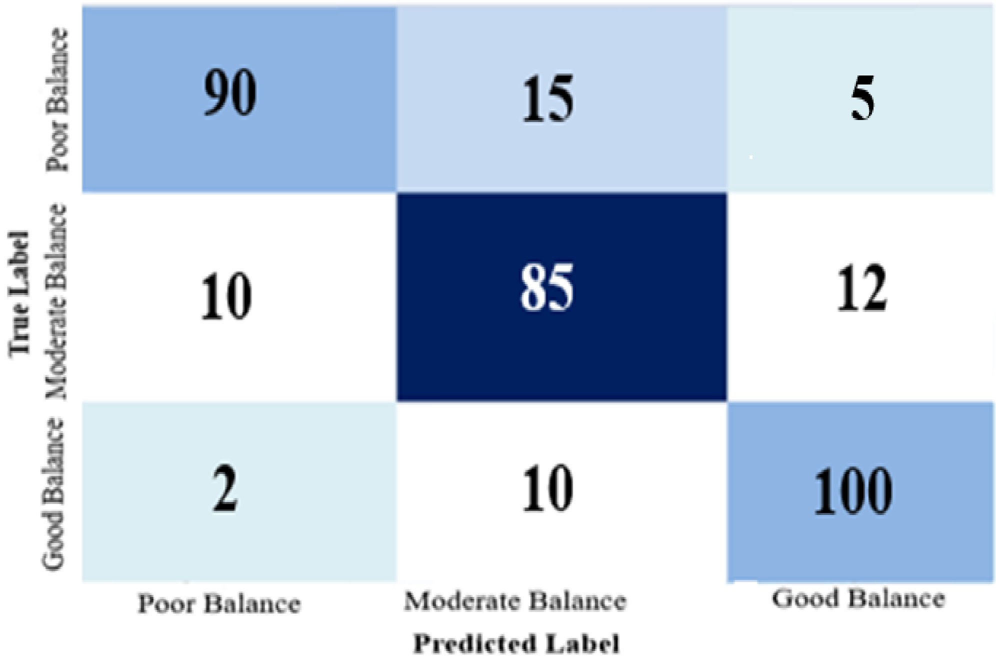

Figure 3. Confusion matrix of SVM

3.3.3 Classification using Navie Bayes (NB)

The Naive Bayes (NB) classifier is an effective and straightforward approach for classification problems, particularly when the characteristics are independent of each other given the class. It exhibits high computational efficiency and effectively handles both categorical and continuous data, making it appropriate for categorizing Work WLB into distinct categories such as "Poor Balance," "Moderate Balance," and "Good Balance.". Bayes' Theorem links the probability of a class based on certain qualities to the prior probability of the class, together with the likelihood of the features given the class. The assumption, which might be regarded as naive, is that all characteristics exhibit conditional independence when the class label is taken into account. Consequently, the probability may be decomposed into constituent components. In order to ascertain the category of a new occurrence, it is necessary to compute the posterior probability for each category and then select the category with the greatest probability. It is then necessary to compute the prior probability for each class. The likelihoods should be estimated, and for continuous features, it is reasonable to assume that they follow a Gaussian distribution. According to figure 4 (and Table 7 – Appendix), 85 occurrences were accurately identified as having Poor Balance. 90 cases accurately classed as Moderate Balance and. 100 cases accurately categorized as Good Balance.

Figure 4. Confusion matrix of NB

4. Results and Discussion

The performance of the three suggested classifiers is evaluated using the following measures: ***Accuracy*** is a quantitative metric that assesses the proportion of properly identified data examples out of the total number of data instances; ***Precision*** is often used in the domains of statistics and machine learning to evaluate the effectiveness of a model, particularly in the realm of binary classification tasks. The statistic measures the proportion of correct positive predictions out of the total number of positive predictions produced by the model; ***Recall***, which is often referred to as sensitivity, is a significant metric used to evaluate performance in binary classification scenarios. The statistic quantifies the ratio of correctly predicted positive cases by the model to the overall number of positive examples in the dataset; ***F1-Score*** is a widely used statistic in the realm of binary classification issues. The statistic integrates both accuracy and recall indicators, providing a comprehensive assessment of a model's effectiveness. The F1 score is particularly valuable in scenarios when the dataset exhibits class imbalance, indicating that one class is disproportionately represented relative to the other class (see Appendix – Table 4).
The categorization of quality metrics is shown in Table 1. The suggested SVM classifier has a 91% accuracy, surpassing the performance of the other two approaches.

Table 1: Performance Comparison of ML classifier with synthetic data generation

Method	Precision	Recall	F1 Score	Accuracy
Random Forest	83	82	82	85
Support Vector Machine	91	94	92	91
Navie Bayes	85	85	83	86

Figure 5 show the performance of the proposed approach, both with and without synthesized data. Based on the outcome, the synthesis data outperforms the genuine data in terms of performance. The classifier categorizes the input data into three classes: bad balanced, moderate balanced, and excellent balanced. Our findings indicate that synthesizing data yields a greater amount of relevant data compared to genuine data, resulting in enhanced classifier performance.

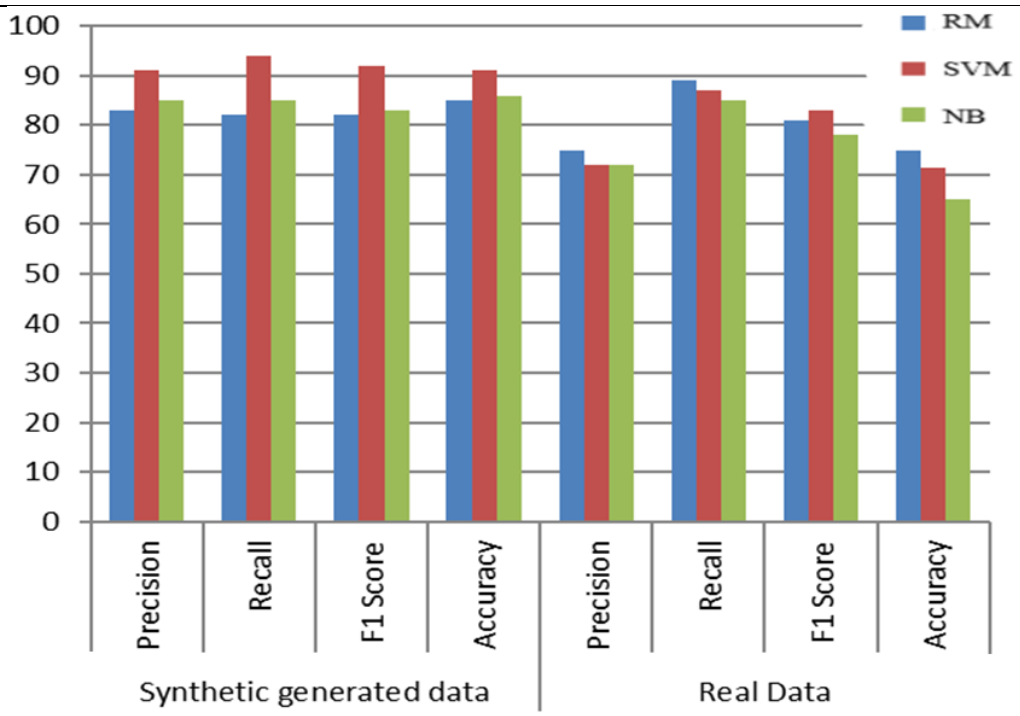

Figure 5: Performance comparison of Proposed Vs Existing

The findings of this study are corroborated by other research conducted under similar conditions. Savage (2023) underscores the necessity of utilising synthetic data in statistical analysis within a comprehensive framework. In order to provide more accurate and controlled training and testing of AI models, it may be possible to create synthetic data in vast quantities at a much faster rate than it would be possible to gather actual data (Dilmegani, 2024). According to Gartner, artificial intelligence (AI) models will heavily rely on synthetic data by 2030. Additionally, 89% of IT executives think that sustaining a competitive edge requires the use of synthetic data (VentureBeat, 2024). In specific contexts, the use of synthetic data can be found above all in search engine companies or e-commerce retailers, less so in other sectors, such as medicine (Walonoski et al, 2018; Arvanitis et al, 2022). Instead, in other sectors, the use of synthetic data is in an embryonic state.

However, existing evaluation frameworks for synthetic data often focus on one or two specific aspects. There is a lack of a comprehensive evaluation framework that merges multiple perspectives and offers a holistic suite of metrics for the assessment of synthetic data. This presents a considerable challenge in the acceptance and reliability of synthetic data generation techniques in relation to these crucial aspects: a. Data fidelity (this aspect concerns the extent to which synthetic data replicates the statistical characteristics of the original dataset); b. Utility (the utility of synthetic data is determined by its effectiveness in facilitating various downstream machine learning tasks). This entails assessing whether models trained on synthetic data demonstrate comparable performance to those trained on real data when evaluated on validation or real-world datasets. Additionally, it is of paramount importance to ensure that the synthetic data does not reveal any sensitive information about individuals in the original dataset, which is a crucial aspect of data privacy. The assessment of privacy can be conducted through the utilisation of techniques such as re-identification risk analysis and the implementation of other privacy-preserving metrics (Shanley et al. 2024; James et al. 2021).

5. Conclusion

This research examines the Work-Life Balance (WLB) of employed women in the Kumbakkonam area by using Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB) classifiers. In order to improve the dataset, synthetic data creation methods were used in conjunction with actual survey data. The results of our study demonstrate that all three classifiers had strong performance when tested with both synthetic and real data. Nevertheless, the RF classifier consistently achieved superior performance compared to SVM and NB in terms of accuracy, precision, and recall. RF demonstrated resilience and efficacy by achieving an accuracy of over 90% with synthetic data and slightly higher with actual data. NB exhibited robust performance, but with significantly less consistency compared to RF. On the other hand, SVM, while effective, demonstrated the lowest performance metrics out of the three. The comparison between synthetic and actual data showed negligible performance disparities, confirming the use of synthetic data for enhancing real-world datasets in WLB investigations. The findings indicate that using sophisticated machine learning methods, namely Support Vector Machines (SVM), may effectively categorize WLB groups. This has the potential to assist in the creation of focused interventions aimed at enhancing the work-life balance of employed women in Kumbakkonam.

References

1. Aafreen, J. H., Balgopal, S., & Vimlesh, T. (2022). An Assessment Women Teachers' Work-Life Balance in Higher Education Institutions, *International Journal For Global Academic & Scientific Research*, 1(4), 17-33.
2. Ahmad, M., & Wani, A.K. (2020). Work-Life Balance of Women in Higher Education Sector: An Assessment of Their Organizational Support, *Journal of International Business and Economy*, 21(1).
3. Aiswarya, V. K. & Velmurugan, V.P. (2022). Work-life Stress and Life Quality of Married Women among Self Financing College Teachers in Kerala. *Webology*. 19(1),
4. Akanji, B., Mordi, C., & Ajonbadi, H. A. (2020). The experiences of work-life balance, stress, and coping lifestyles of female professionals: insights from a developing country, *Employee Relations*, 42(4), 999-1015. https://doi.org/10.1108/ER-01-2019-0089
5. Arvanitis, T.N., White, S., Harrison, S., Chaplin, R., Despotou, G.(2022). A method for machine learning generation of realistic synthetic datasets for validating healthcare applications. *Health Informatics Journal,* 28(2). doi:10.1177/14604582221077000.
6. Bharadwaj, V. K., & Shanker, M. (2021, January 15). Attaining Work-Life Balance is a Pressing Priority: A Theoretical Review. ANVESHAK-International Journal of Management, 9(2), 9. https://doi.org/10.15410/aijm/2020/v9i2/155634.
7. Burke, R.J. (2001). Organizational values, work experiences and satisfactions among managerial and professional women. *Journal of Management Development,* 20(4), 346-354. https://doi.org/10.1108/02621710110695333
8. Dilmegani, C. (2024). Synthetic data vs real data: Benefits, challenges in 2024. https://research.aimultiple.com/synthetic-data-vs-real-data/, 2024. Accessed: 2024-09-19
9. Figueira, A., & Vaz, B. (2022). Survey on Synthetic Data Generation, Evaluation Methods and GANs, *Mathematics,* 10(15), 2733. https://doi.org/10.3390/math10152733.
10. Fonseca, J., & Bacao, F.(2023). Tabular and latent space synthetic data generation: a literature review, *Journal of Big Data,* 10(115). https://doi.org/10.1186/s40537-023-00792-7.
11. James, S., Harbron, C., Branson, J., Sundler, M. (2021). Synthetic data use: exploring use cases to optimise data utility. *Discover Artificial Intelligence,* 1-15. https://doi.org/10.1007/s44163-021-00016-y
12. Lekchiri, S., & Eversole, B.A.W. (2020). Perceived work-life balance: Exploring the experiences of professional Moroccan women, *Human Resource Development Quarterly*, 32(1), 35–53. https://doi.org/10.1002/hrdq.21407
13. Muafi, M., & Marseno, W.A. (2021). The effects of work-life balance and emotional intelligence on organizational commitment mediated by work engagement. *International Journal of Business Ecosystem & Strategy,* 3(2), 1 – 15.
14. Noronha, S.D., & Aithal, P. (2019a). Influence of socio-Demographic factors on Work-life balance (WLB) among Public Universities teachers in Karnataka, *International Journal of Scientific Research*, 8(11): 2277- 2281.
15. Noronha, S.D., & Aithal, P.S. (2019b). Work-Life balance among Women Employees: A study on initiatives undertaken by Indian Organizations, *MERC Global's International Journal of Management*, 7(1): 22-27.
16. Noronha, S.D., & Aithal, P.S. (2020). Work-Life Balance Issues Encountered by Working Women in Higher Education Sector. *Scholedge International Journal of Management & Development*, 7(5), 72. https://doi.org/10.19085/sijmd070501.
17. Pace, F.; Sciotto, G. (2022). Gender Differences in the Relationship between Work–Life Balance, Career Opportunities and General Health Perception, *Sustainability*, 14, 357. https://doi.org/10.3390/su14010357.
18. Pandey, N. (2023). Exploring the Factors Affecting the Work-Life Balance of ASHA's Health Workers and their Impact on Performance in Rural Uttarakhand, *International Journal of Scientific Research in Engineering and Management*, 07(4), 1 – 10. https://doi.org/10.55041/ijsrem20210
19. Pandey, N., & Dhanopia, M. (2019). Work-Life balance and working Indian mothers: An empirical study, *International Journal of Engineering Applied Science and Technology*, 4(7), 119–124. https://doi.org/10.33564/ijeast.2019.v04i07.018 Pandey, N. (2023).
20. Popoola, S.O., & Fagbola, O.O. (2020). Work-Life Balance, Self-Esteem, Work Motivation, and Organizational Commitment of Library Personnel in Federal Universities in Southern Nigeria, *International Information & Library Review, 53*, 214 - 228.
21. Reddy, N.K., Vranda, M.N., Ahmed, A., Nirmala, B.P., & Siddaramu, B. (2010). Work–Life Balance among Married Women Employees, *Indian Journal of Psychological Medicine,* 32(2), 112 - 118.
22. Savage, N. (2023). Synthetic data could be better than real data. *Nature*. doi: 10.1038/d41586-023-01445-8.

23. Sethi, U.J. (2014). Influence of Work Life Balance on Organisational Commitment: A Comparative Study of Women Employees Working in Public and Private Sector Banks. *European Journal of Business and Management,* 6(34), 215-219.
24. Shabir, S., Khan, O. F., & Gani, A. (2021). Work-life interference: a perpetual struggle for women employees, *International Journal of Organizational Analysis*, 30(2), 181–196. https://doi.org/10.1108/ijoa-04-2020- 2133.
25. Shaikh, N., Chandio, J.A.(2024). Work-Life Balance Policies and Organizational Outcomes: A Qualitative Investigation, *Bulletin of Business and Economics (BBE)*, 13(1). https://doi.org/10.61506/01.00230.
26. Shanley, D., Hogenboom, J., Lysen, F., Wee, L., Lobo, G. A., Dekker, A., Meacham, D. (2024). Getting real about synthetic data ethics: Are AI ethics principles a good starting point for synthetic data ethics? *EMBO Rep.*, 25(5):2152-2155. doi: 10.1038/s44319-024-00101-0.
27. Thi, T.D., Ngo, A.T., Duong, N.T., & Pham, V.K. (2021). The Influence of Organizational Culture on Employees' Satisfaction and Commitment in SMEs: A Case Study in Vietnam, *The Journal of Asian Finance, Economics and Business*, 8(5), 1031–1038. https://doi.org/10.13106/JAFEB.2021.VOL8.NO5.1031
28. Vasumathi, A.(2018). Work life balance of women employees: a literature review, *Inernational. Journal of Services and Operations Management*, 29,(1), 100 – 146.
29. VentureBeat (2024). 89% of tech execs see synthetic data as a key to staying ahead. https://venturebeat.com/ai/89-of-tech-execs-see-synthetic-data-as-a-key-to-staying-ahead, 2021. Accessed: 2024-09-19.
30. Walonoski, J., Kramer. M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Dube, K., Gallagher, T., McLachlan, S. (2018). Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record, *Journal of the American Medical Informatics Association,* 1;25(3):230-238. doi: 10.1093/jamia/ocx079.

Appendix

Table 1: Sample data form input dataset

Age	Marital Status	No of Children	Job Satisfaction	Stress Level
35	Married	2	High	Medium
42	Single	0	Medium	Low
29	Divorced	1	Low	High
54	Widowed	3	Medium	Medium
31	Married	1	High	Low
NaN	Single	NaN	High	Medium
45	Married	2	Low	High
38	Divorced	0	Medium	Low
50	Widowed	3	High	Medium
28	Married	1	Low	High

Table 2: Missing Value imputed

Age	Marital Status	No of Children	Job Satisfaction	Stress Level
35	Married	2	High	Medium
42	Single	0	Medium	Low
29	Divorced	1	Low	High
54	Widowed	3	Medium	Medium
31	Married	1	High	Low
39	Single	NaN	High	Medium
45	Married	2	Low	High
38	Divorced	0	Medium	Low
50	Widowed	3	High	Medium
28	Married	1	Low	High

Table 3: Values after one-hot encoding

Age	No of children	Married	Single	Divorced	Widowed	High_JS
35	2	1	0	0	0	1
42	0	0	1	0	0	0
29	1	0	0	1	0	0
54	3	0	0	0	1	0
31	1	1	0	0	0	1
39	1	1	0	0	0	1
45	2	1	0	0	0	0
38	0	0	0	1	0	0
50	3	0	0	0	1	1
28	1	1	0	0	0	0

Table 4: The performance measures

Accuracy	$Accuracy = \frac{TrueNegative+TruePositive}{Truepositive+FalsePositive+TrueNegative+FalseNegative}$
Precision	$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$
Recall	$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$
F1 - score	$F1 - Score = 2\frac{(Precision\ x\ Recall)}{(Precision + Recall)}$

Table 5: RF Classification report

	precision	recall	f1-score	support
0	0.89	0.87	0.88	100
1	0.83	0.82	0.82	100
Accuracy			0.85	100
macro-avg	0.85	0.85	0.85	300
weighed avg	0.85	0.85	0.85	300

Table 6: SVM Classifier Classification report

	precision	recall	f1-score	support
0	0.92	0.91	0.91	105
1	0.91	0.94	0.92	103
Accuracy			0.91	315
macro-avg	0.91	0.91	0.91	315
weighed avg	0.91	0.91	0.91	315

Table 7: Classification report of Navie Bayes

	precision	recall	f1-score	support
0	0.88	0.85	0.87	110
1	0.85	0.85	0.83	110
Accuracy			0.86	330
macro-avg	0.85	0.86	0.86	330
weighed avg	0.85	0.86	0.85	330