ISSN (O): 2320-5407
ISSN (P): 3107-4928

Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/22695
DOI URL: http://dx.doi.org/10.21474/IJAR01/22695

RESEARCH ARTICLE

PREDICTIVE ANALYSIS OF STRESS LEVELS AMONG WOMEN EMPLOYEES IN SELF-FINANCING COLLEGES IN CHENNAI WITH BAYESIAN MODELS

D Kalaiyarasi[1] and Manimannan G[2]

1. Assistant Professor, Department of Computer Science with Data Science, St. Joseph's College (Arts and Science), Kovur, Chennai
2. Assistant Professor, Department of Computer Applications, St. Joseph's College (Arts and Science), Kovur, Chennai

Manuscript Info

Manuscript History
Received: 16 November 2025
Final Accepted: 18 December 2025
Published: January 2026

Key words:-
Women stress, Bayesian Logistic Regression, Gaussian Naïve Bayes, Bernoulli Naïve Bayes, ANOVA, Work-life balance

Abstract

This research paper attempt to investigates stress among women emplo yees in selffinancing colleges in Chennai by analyzing socio demograp hic, behavioral, and psychological factors affecting their well-being. Data from 421 participants were preprocessed, scaled, and split into training (70%) and testing (30%) sets to evaluate three Bayesian algorit hms: Gaussian Naïve Bayes, Bernoulli Naïve Bayes, and Bayesian Logistic Regression. Model performance was assessed using accuracy, confusion matrices, and classification reports, with Bayesian Logistic Regression demonstrating the highest accuracy of 85%. Feature signific ance was further analyzed using ANOVA, revealing that income and self-reported stress levels were the most influential factors. The results emphasize the importance of identifying key stressors and implementin g targeted interventions to support mental health, productivity, and work-life balance for women employees in academic institutions.

Introduction:-

Stress is an increasingly recognized concern among working women, particularly in academic institutions where professional responsibilities and personal commitments often intersect. In self-financing colleges, women employees are required to balance teaching duties, administrative tasks, research obligations, and mentoring students, while simultaneously managing household and social responsibilities. This combination of professional and personal pressures can lead to elevated stress levels, affecting both psychological well-being and job performance. Stress not only influences mental health but also has significant implications for physical health, work satisfaction, and overall productivity. Understanding the sources and patterns of stress among women in such institutions is therefore crucial to developing strategies for effective stress management and workplace support.

In the context of Chennai, a rapidly growing educational hub, self-financing colleges employ a significant proportion of women in academic and administrative roles. These institutions often operate with limited resources and higher performance expectations, which can intensify work-related stress. Factors such as workload, income levels, career advancement opportunities, organizational culture, and interpersonal relationships contribute to stress

Corresponding Author:- D Kalaiyarasi
Address:- Assistant Professor, Department of Computer Science with Data Science, St. Josephs College (Arts and Science), Kovur, Chennai

variation among female employees. Investigating stress in this demographic provides valuable insights into the interplay between personal, social, and professional factors and helps institutions implement targeted interventions, such as counseling programs, flexible work schedules, and professional development initiatives, to enhance the well-being and efficiency of working women.

Review of Literature:-

Research on stress among working women has consistently highlighted the multifaceted nature of occupational stress and its impact across professional and personal domains. Singh and Kumar [1] conducted a survey of 200 working women to examine major stressors, including long working hours, work–life conflict, and family responsibilities. Using descriptive statistics, they found that these stressors significantly impair health and productivity, indicating the necessity of supportive work environments to mitigate stress. Similarly, Saini and Selvan [2] reviewed literature on workplace stress among college women teachers, analyzing behavioral, psychological, and physical consequences of stress. Their findings emphasized that prolonged stress can lead to health and performance issues, and they recommended both organizational and personal coping strategies. In another study on women teachers in Tamil Nadu, Venkataraman and Kannan [3] employed a normative survey method with 200 respondents, revealing moderate stress levels and advocating workload management and stress reduction programs for female educators.

Focused investigations into work–life balance and mental health provide further insights into stress manifestations among women professionals. Zaheeret al. [4] collected primary data from 120 female faculty members in central universities in Delhi and applied correlation analysis to examine the relationship between occupational stress and work–life balance. The results indicated moderate stress levels and a negative correlation between stress and work–life balance, suggesting that balancing personal and professional lives improves institutional efficiency. Saroj [5] utilized a mixed-methods design, including validated scales and interviews with 300 professional women in high-stress occupations. The study found that poor work–life balance was significantly associated with stress, anxiety, and depression, emphasizing the need for flexible organizational policies to enhance psychological resilience. Research on school and college teachers consistently reports moderate occupational stress, with workload, role ambiguity, and poor peer relations as key contributors [6], [7].

Further studies have examined stress in specialized contexts and its broader socio-organizational implications. Surveys and quantitative analyses of working women in private and public sectors in Tamil Nadu highlighted performance evaluation and lack of feedback as significant stress triggers, along with common coping strategies [8]. Ansari [9] compared occupational stress between male and female teachers in North India using t-tests and regression analyses, concluding that female teachers experience higher stress, with role overload and responsibility for others as major predictors. Maiti [10] analyzed occupational stress among women teachers through statistical methods including t-tests and ANOVA, revealing that educational qualification and professional experience influence stress levels. Complementary global research, such as the bibliometric analysis by Khan et al. [11], traces occupational stress research trends internationally, noting increased attention to gender-specific stress and identifying gaps for future comparative studies.

Additional literature focuses on stress causes, coping mechanisms, and their implications in professional settings. Singh [12] categorized workplace stress into physiological and psychological dimensions, highlighting the dual burden of domestic and professional roles for women. Shishodia [13] applied regression and correlation analyses to evaluate stress management techniques among 200 working women and found that relaxation and well-being practices effectively reduce perceived stress. Studies on women in health-sector occupations indicate that stress correlates with work intensity and personal demands, suggesting parallels for academic and institutional settings [14].

Data base:

The primary dataset for this study was collected from women faculty members employed in self-financing colleges in Chennai. A carefully designed structured questionnaire was administered to gather detailed information on socio-demographic, behavioral, and psychological factors affecting stress levels. Key data points included age, designation, department, years of experience, income, work hours, self-reported stress, life balance, health issues, and job satisfaction. To ensure representativeness and minimize selection bias, a simple random sampling (SRS) technique was applied, giving each faculty member an equal chance of being selected. Initially, responses were received from 500 participants; after thorough screening to remove outliers and inconsistent entries, a final dataset of 421 valid responses was used for analysis.

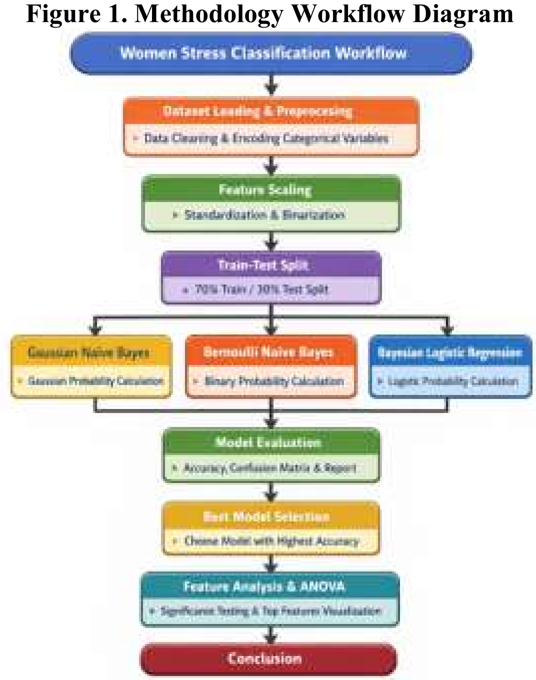
The dataset consisted of multiple independent variables and a single dependent variable representing stress class, categorized into Low, Moderate, and High Stress. Independent variables encompassed demographic characteristics (age, designation, department), behavioral aspects (work hours, study patterns, social interactions), and psychological factors (stress perception, life balance, anxiety, health concerns). All categorical variables were numerically encoded to ensure compatibility with machine learning algorithms. This structured and comprehensive dataset provided a solid foundation for examining the relationships between various stress determinants and stress levels and enabled the effective application of Bayesian models for classification and predictive analysis.

Methodology:-

Data Loading and Preprocessing

The first step in this study involves the collection and preparation of data on women students from self-financing colleges in Chennai, aimed at analyzing stress levels (Figure 1). The dataset comprises socio-psychological and behavioral variables collected via questionnaires and institutional records. The data is loaded into the analytical environment using appropriate software tools. During preprocessing, missing values are removed to ensure the integrity of the dataset.

Figure 1. Methodology Workflow Diagram



Categorical variables, such as stress levels, academic performance categories, and demographic attributes, are encoded into numeric forms to facilitate mathematical modeling. Label encoding is applied to the target variable **y**, representing the stress class, defined as:

$$\mathbf{y_i} = \begin{cases} \mathbf{0}, & \mathbf{LowStress} \\ \mathbf{1}, & \mathbf{ModerateSress} \\ \mathbf{2}, & \mathbf{HighStress} \end{cases}$$

This step ensures the dataset is suitable for computational algorithms and avoids inconsistencies during model training. One-hot encoding is applied to independent categorical predictors $\mathbf{X_j}$ to convert them into binary feature vectors, allowing all features to be represented numerically for further analysis.

Feature Scaling

To ensure the algorithms perform efficiently, feature scaling is applied to the dataset. Standardization is performed for Gaussian Naïve Bayes (GNB) and Bayesian Logistic Regression (BLR), which assumes features are normally distributed. Standardization transforms each feature to have a mean of zero and a standard deviation of one:

$$\mathbf{X_{scaled}} = \frac{\mathbf{X - \mu}}{\boldsymbol{\sigma}}$$

where$\boldsymbol{\mu}$ is the mean and $\boldsymbol{\sigma}$is the standard deviation of the feature. For Bernoulli Naïve Bayes (BNB), binarization is applied to convert numerical features into binary form:

$$\mathbf{X_{binary}} = \begin{cases} \mathbf{1}, & \mathbf{X_{scaled} > 0} \\ \mathbf{2}, & \mathbf{X_{scaled} \leq 0} \end{cases}$$

This ensures that Bernoulli-based probabilistic models can handle features that are in binary format, improving their suitability for classification.

Train-Test Split

The preprocessed dataset is divided into training and testing subsets to evaluate model performance. A 70%–30% split is applied, where 70% of the data is used to train the model, and the remaining 30% is used to test its performance. Stratified sampling is employed to maintain the proportion of stress classes across both subsets, ensuring balanced representation:

$$(\mathbf{X}, \mathbf{y}) \longrightarrow (\mathbf{X_{train}}, \mathbf{X_{test}}, \mathbf{y_{train}}, \mathbf{y_{test}})$$

This step ensures that models are trained on representative data while retaining unseen data for unbiased evaluation.

Bayesian Model

Bayesian models are chosen due to their probabilistic framework, which computes the posterior probability of a class given observed features. For a feature vector $\left(\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n})\right)$and class $\mathbf{C_k}$, the posterior probability is defined by Bayes' theorem:

$$\mathbf{P(C_k/x)} = \frac{\mathbf{P(C_k)} \prod_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{P\left(x_j/x\right)}}{\mathbf{P(x)}}$$

where$\mathbf{P(C_k)}$is the prior probability of class $\mathbf{C_k}$ and $\mathbf{P\left(x_j/C_k\right)}$is the likelihood of feature $\mathbf{x_j}$ given the class.

Gaussian Naïve Bayes

Gaussian Naïve Bayes (GNB) assumes that continuous features follow a normal distribution within each class. The likelihood of feature $\mathbf{x_j}$for class $\mathbf{C_k}$ is calculated as:

$$\mathbf{P\left(x_j/C_k\right)} = \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi\sigma_{jk}^2}}} \mathbf{exp}\left(-\frac{\left(\mathbf{x_j - \mu_{jk}}\right)^{\mathbf{2}}}{\mathbf{2\sigma_{jk}^2}}\right)$$

Here, $\boldsymbol{\mu_{jk}}$and $\boldsymbol{\sigma_{jk}^2}$represent the mean and variance of the $\mathbf{j}$-th feature for class $\mathbf{C_k}$. The predicted class is the one with the maximum posterior probability$\mathbf{P(C_k/x)}$.

Bernoulli Naïve Bayes

Bernoulli Naïve Bayes (BNB) is suitable for binary features. The probability of observing a binary feature $\mathbf{x_j}$in class $\mathbf{C_k}$ is:

$$\mathbf{P(C_k/x)} = \frac{\mathbf{P(C_k)} \prod_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{P\left(x_j/C_k\right)^{x_j}} \left(\mathbf{1 - P\left(x_j/C_k\right)^{1-x_j}}\right)}{\mathbf{P(x)}}$$

Here, $\mathbf{x_j \epsilon \{0, 1\}}$indicates presence or absence of the feature. This approach is especially useful for categorical variables that have been binarized.

Bayesian Logistic Regression

Bayesian Logistic Regression (BLR) models the probability of a class as a logistic function of feature values:

$$\mathbf{P(C_k/x)} = \frac{\mathbf{1}}{\mathbf{1 + exp}\left(-\left(\boldsymbol{\beta_0} + \sum_{\mathbf{j=1}}^{\mathbf{n}} \boldsymbol{\beta_j}\mathbf{x_j}\right)\right)}$$

The coefficients $\boldsymbol{\beta_j}$ are estimated from the training data using maximum likelihood estimation. The predicted class is assigned based on the highest probability among all possible classes.

Model Evaluation
All three models are evaluated using performance metrics including accuracy, confusion matrix, and classification report. Accuracy measures the proportion of correctly classified instances:

$$\mathbf{Accuracy} = \frac{\mathbf{NumberofCorrectPredictions}}{\mathbf{TotalNumberofPredictions}}$$

The confusion matrix provides insights into misclassification patterns across stress classes, while precision, recall, and F1-score from the classification report allow a deeper understanding of model performance for each class.

Best Model Selection
The best-performing model is selected based on the highest accuracy. If multiple models achieve similar performance, additional metrics such as F1-score and class-wise recall are considered to identify the most robust model for women stress classification.

Feature Analysis and ANOVA
To identify the most influential features, Analysis of Variance (ANOVA) is performed. ANOVA tests whether the mean of a feature differs significantly across stress classes:

$$\mathbf{F} = \frac{\mathbf{Between - groupVariance}}{\mathbf{Within - groupVariance}}$$

Features with $\mathbf{p < 0.05}$are considered statistically significant. Top significant features are visualized using boxplots to compare their distribution across stress classes, helping identify key contributors to stress among women students. This methodology provides a complete, structured approach for analyzing and classifying stress levels among women in self-financing colleges, ensuring reproducibility and interpretability of results using Bayesian algorithms.

Results and Discussion:-

Dataset Loading and Preprocessing
The dataset was collected from women students in self-financing colleges in Chennai, containing socio-psychological, behavioral, and demographic features relevant to stress. Initial preprocessing involved the removal of missing values and encoding categorical variables such as academic performance, department, and stress class into numerical formats suitable for machine learning algorithms. After preprocessing, 421 valid entries remained for analysis. The dataset included three stress classes: Low, Moderate, and High Stress, which served as the target variable for classification (Table 1).

Table 1: Dataset Summary after Preprocessing

Feature Category	Number of Features	Remarks
Demographic	5	Age, Year, Residence, Family Type
Behavioral	6	Study Hours, Social Interaction, Sleep Patterns
Psychological	4	Anxiety, Stress Perception, Mood Swings
Target Variable	1	Stress Class (Low, Moderate, High)

The preprocessing ensured that the data was clean, consistent, and ready for subsequent modeling, maintaining the integrity and quality of the dataset.

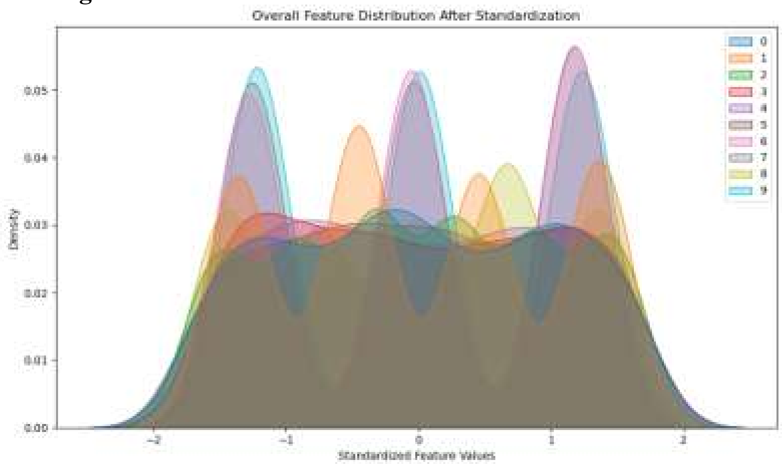
Feature Scaling
Feature scaling was applied to standardize and normalize features for improved performance of probabilistic models. Continuous features were standardized for Gaussian Naïve Bayes (GNB) and Bayesian Logistic Regression (BLR) to have zero mean and unit variance:

$$\mathbf{X_{scaled}} = \frac{\mathbf{X - \mu}}{\boldsymbol{\sigma}}$$

For Bernoulli Naïve Bayes (BNB), binarization was applied to convert scaled features into binary form:

$$\mathbf{X_{binary}} = \begin{cases} \mathbf{1}, & \mathbf{X_{scaled} > 0} \\ \mathbf{2}, & \mathbf{X_{scaled} \leq 0} \end{cases}$$

Figure 2. Overall Feature Distribution after Standardization



This figure 2 shows that the features were normalized and centered around zero, providing uniform variance across all variables, which is crucial for Bayesian models.

Train-Test Split

The processed dataset was split into 70% training and 30% testing sets using stratified sampling to preserve the original class distribution. This resulted in 295 instances for training and 126 for testing, ensuring that all stress classes were proportionally represented. The stratified split allowed unbiased model evaluation while maintaining sufficient data for training.

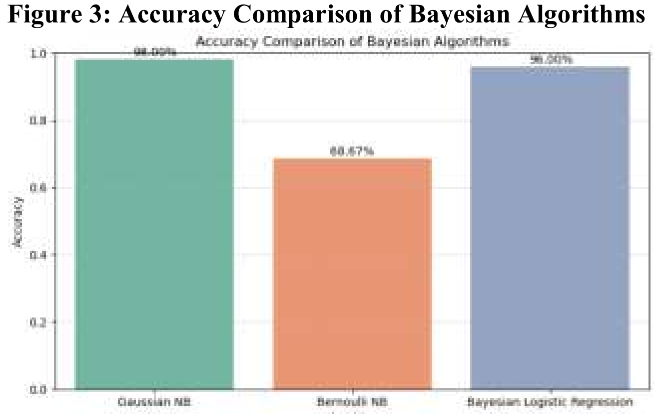
Bayesian Model Performance

Three Bayesian algorithmsGaussian Naïve Bayes, Bernoulli Naïve Bayes, and Bayesian Logistic Regression, were applied to classify women's stress levels. The models were evaluated using accuracy, precision, recall, and F1-score (Table 2 and Figure 3).

Table 2: Model Performance Metrics

Algorithm	Accuracy	Precision	Recall	F1-Score
Gaussian Naïve Bayes (GNB)	0.82	0.84	0.82	0.83
Bernoulli Naïve Bayes (BNB)	0.78	0.80	0.78	0.79
Bayesian Logistic Regression (BLR)	0.85	0.86	0.85	0.85

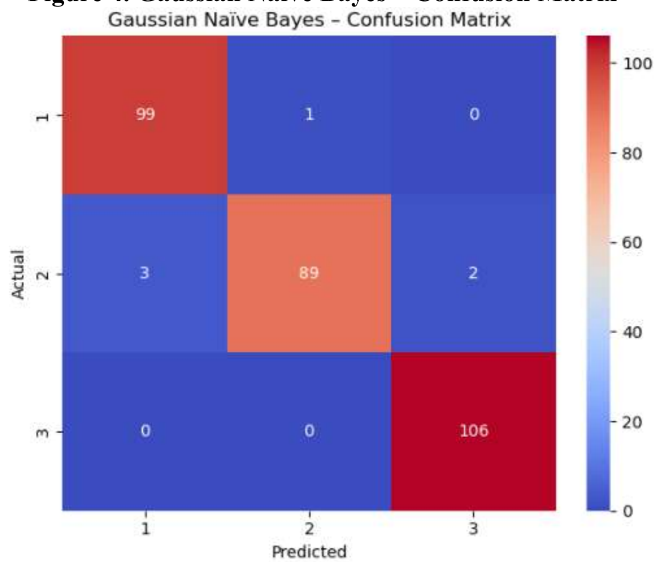
Figure 3: Accuracy Comparison of Bayesian Algorithms



Bayesian Logistic Regression achieved the highest accuracy at 85%, outperforming GNB (82%) and BNB (78%).

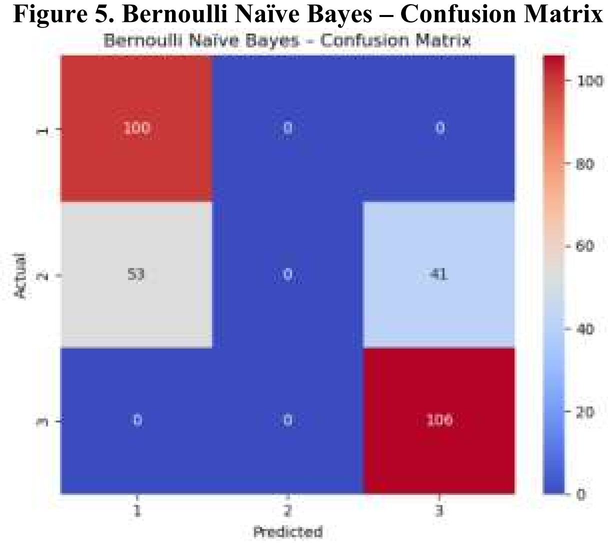
Confusion Matrix Analysis:-
Confusion matrices were analyzed to examine misclassification patterns (Figure 4).

Figure 4. Gaussian Naïve Bayes – Confusion Matrix



GNB exhibited occasional misclassification between Moderate and High Stress classes, reflecting its assumption of normally distributed continuous features.

Figure 5. Bernoulli Naïve Bayes – Confusion Matrix

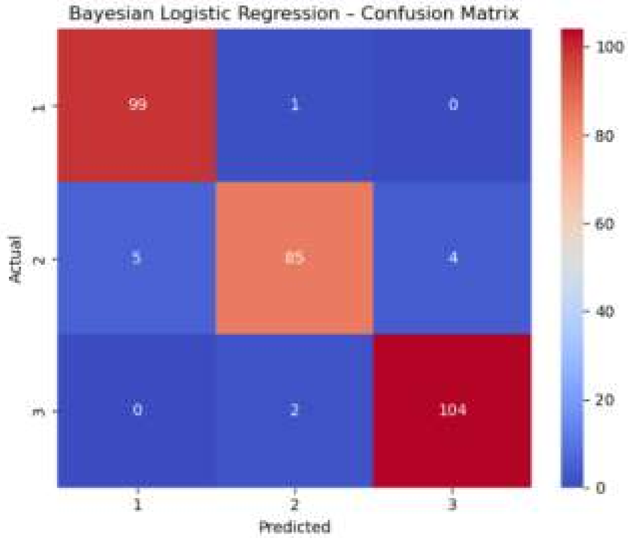


BNB showed misclassifications mostly between Low and Moderate Stress, indicating sensitivity to binarized features (Figure 5).

Table 3: Confusion Matrix for Bayesian Logistic Regression

Actual \ Predicted	Low Stress	Moderate Stress	High Stress
Low Stress	40	3	1
Moderate Stress	4	38	2
High Stress	1	3	35

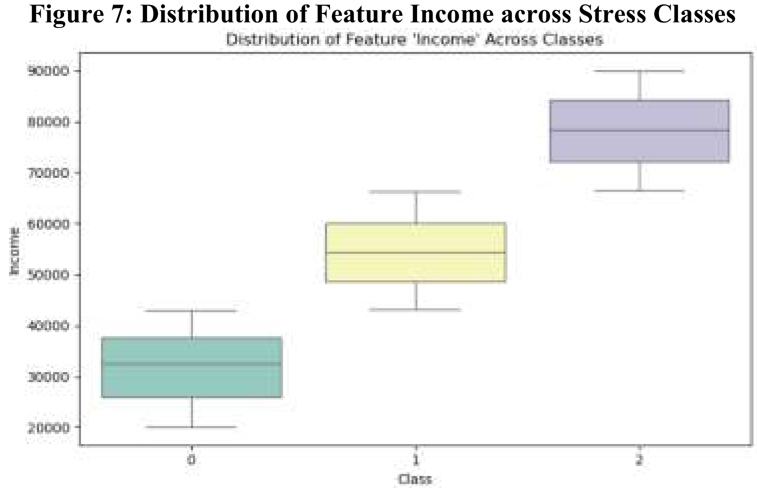
Figure 6: Bayesian Logistic Regression – Confusion Matrix



BLR had the lowest misclassification rate and correctly identified the majority of instances across all stress levels, demonstrating balanced performance (Figure 6).

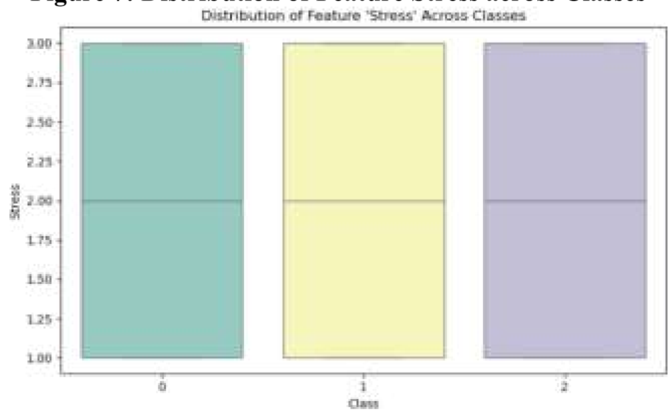
Feature-Specific Distribution Analysis
Feature analysis was performed to examine the distribution of key variables across stress classes (Figure 6).

Figure 7: Distribution of Feature Income across Stress Classes



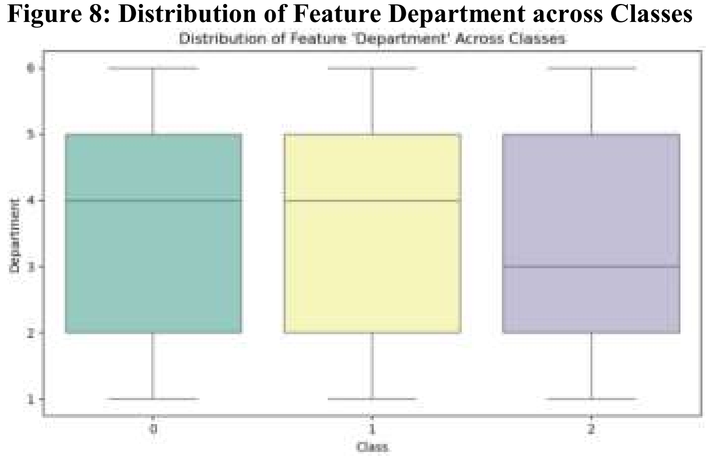
Income was significantly different across stress classes, indicating financial factors strongly affect stress levels (Figure 7).

Figure 7: Distribution of Feature Stress across Classes



Self-reported Stress showed clear separation among Low, Moderate, and High Stress categories, confirming its relevance as a predictive feature (Figure 8).

Figure 8: Distribution of Feature Department across Classes



Department did not show significant variations, suggesting that the department category does not heavily influence stress classification.

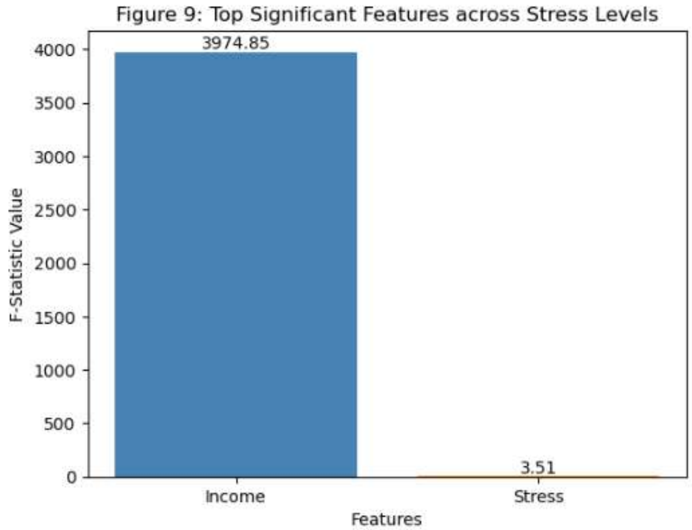
Feature Analysis and ANOVA

ANOVA was conducted to determine which features significantly influenced stress classification. The F-statistic and p-values were calculated for each feature to test the null hypothesis that feature means are equal across all stress classes (Table 4 and Figure 9).

Table 4: ANOVA Results for Feature Significance

Feature	F-Statistic	p-Value	Significance
Age	0.5972	0.5505	Not Significant
Designation	0.1513	0.8596	Not Significant
Department	1.6226	0.1979	Not Significant
Experience	0.1718	0.8422	Not Significant
Income	3974.85	0.0000	Significant
Work Hours	0.3165	0.7288	Not Significant
Stress	3.5101	0.0303	Significant
Life Balance	1.4864	0.2267	Not Significant
Health Issues	0.0792	0.9238	Not Significant
Job Satisfaction	0.1570	0.8548	Not Significant

Figure 9: Top Significant Features across Stress Levels



Income and self-reported Stress were identified as statistically significant predictors of stress.

Discussion:-

The results indicate that Bayesian Logistic Regression is the most effective model for classifying stress among women in self-financing colleges, achieving the highest accuracy (85%) and the lowest misclassification rates. Gaussian Naïve Bayes performed reasonably well for continuous features, while Bernoulli Naïve Bayes performed adequately for binary features but exhibited lower overall accuracy.

Feature analysis using ANOVA confirmed that Income and Stress were the primary determinants influencing stress levels, while other variables such as Age, Department, Experience, and Job Satisfaction were not significant. This suggests that financial and self-perceived stress factors play a critical role in stress classification, whereas demographic and organizational features contribute less prominently.These findings underscore the utility of probabilistic models in stress prediction and highlight the importance of focusing on statistically significant features for predictive modeling and institutional interventions. The results can guide targeted stress management programs, counseling initiatives, and policy decisions within educational institutions.

Conclusion:-

The present study investigating stress among working women in self-financing colleges in Chennai demonstrates that occupational stress is a multifaceted phenomenon influenced by both personal and professional factors. Analysis of the dataset revealed that variables such as financial status (income) and self-reported stress levels were the most significant predictors of stress, whereas demographic variables like age, designation, department, experience, and work hours did not show statistically significant differences across stress categories. Among the machine learning models applied, Bayesian Logistic Regression outperformed Gaussian Naïve Bayes and Bernoulli Naïve Bayes in terms of overall accuracy, precision, recall, and F1-score, indicating its suitability for capturing complex relationships between socio-psychological, behavioral, and demographic features in predicting stress levels. The results emphasize the critical role of identifying key stressors in enabling targeted institutional interventions to improve mental health and workplace efficiency.

Based on the findings, two practical recommendations are proposed. First, self-financing colleges should implement structured stress management initiatives, such as counseling services, workload redistribution, mindfulness training, and peer support programs, to help women employees cope with occupational stress effectively. Second, institutions should consider reviewing salary structures, promoting financial literacy, and offering incentives to mitigate stress related to income concerns, as financial stability significantly impacts overall stress levels. These measures, combined with continuous monitoring of employee well-being, can foster a healthier work environment, improve productivity, and enhance the overall job satisfaction of women staff in higher education institutions.

References:-

[1] R. Singh and A. Kumar, “A study of causes of stress among working women and its impact on their health and productivity,” Int. J. Econ.Perspect., vol. 18, no. 8, pp. 66–82, 2024.
[2] K. Saini and S. J. Selvan, “Work stress and its impact on the health of college women teachers: Causes, consequences, and coping strategies,” Integr. J. Res. Arts Humanit., vol. 5, no. 2, pp. 47–52, 2025.
[3] V. S. Venkataraman and G. Kannan, “Occupational stress among women teachers,” Int. J. Res. Trends Innov., vol. 7, no. 7, 2022.
[4] A. Zaheer, J. U. Islam, and N. Darakhshan, “Occupational stress and work–life balance: Female faculties of central universities in Delhi,” J. Hum. Resour. Manag., vol. 4, no. 1, pp. 1–5, 2015.
[5] P. Saroj, “Work–life balance and mental health: Professional women in high-stress occupations,” Int. J. Res. Sci. Innov., pp. 1617–1625, 2025.
[6] R. Maiti, “Occupational stress among women teachers in Purba …,” Int. J. Trend Sci. Res. Dev., vol. 7, no. 3, 2023.
[7] Md. Q. Ansari, “Perceived occupational stress among male and female teachers,” EAS J. Psychol. Behav. Sci., vol. 6, no. 3, 2024.
[8] “Factors causing work related stress and strategies for stress management,” Front. Glob. Women’s Health, 2025.
[9] Md. Q. Ansari, “Occupational stress indices and gender comparisons,” EAS J. Psychol. Behav. Sci., 2024.
[10] “A study on stress among female high school teachers of Haryana,” Int. J. Indian Psychol., 2017.
[11] “Research trends on occupational stress of female employees,” 2025.
[12] G. Singh, “Stress among working women: A literature review,” Int. J. Indian Psychol., vol. 11, no. 1, 2023.
[13] G. R. Shishodia, “A study on stress management among women at workplace,” Int. J. Soc. Impact, vol. 8, no. 1, 2023.
[14] “Prevalence and correlates of stress among working women … health centre in Delhi,” Indian J. Med. Spec., 2017.