SUPPORT VECTOR MACHINE BASED PREDICTION OF WORK LIFE BALANCE AMONG WOMEN IN INFORMATION TECHNOLOGY ORGANIZATION

Emidi Anusha[1], Dr.Ch.V.Phani Krishna[2]

[1] PG Scholars, Department of CSE, Teegala Krishna Reddy Engineering College**,**Hyderabad, Telangana, India.

[2] Professor, Department of CSE, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana, India.

ABSTRACT:

This article employs supervised machine learning (ML) to develop a prediction model for the degree of work-life balance of information technology (IT) women professionals. In total, 425 women in the IT industry took part in this article. The data needed is based on a questionnaire with 40 parameters, which serves as the input for supervised ML.Different ML models are created and compared, such as regression trees and support vector machines (SVM). The coefficient of determination (R2) are used to assess the performance of ML models. For work-life balance prediction, an efficient SVM model is found to have the highest accuracy. A comparison of the prediction capabilities of an SVM model and a multiple regression model is also performed. When the accuracy of the two prediction models is compared, SVM outperforms multiple regression. SVM is presented as an alternative prediction modelling technique to multiple regression in organisational behaviour studies in this article. Step-by-step instructions for ML modelling, optimization, training, validation, and result interpretation are provided to researchers interested in ML modelling.

Keywords: Machine Learning, Random Forest, Logistic Regression, Decision Tree, SVM and ML techniques, evaluation.

DOI: 10.48047/ecb/2023.12.Si8.681

INTRODUCTION

THE information technology (IT) industry in India is the largest source of technical manpower in the world. It is the second-largest employer of women in India [1]. The IT industry has provided women with remarkable opportunities in terms of career and

stimulating environment [2], [3]. However, the industry is demanding in terms of extended work hours, up skilling with technology, overseas client calls at odd hours, stringent deadlines, etc. [4]. According to the report, Women and IT Scorecard - India (2017), women form 51% of the workforce at the entry-level, 25% form the managerial positions, however, only 1% reach the Chief Executive positions [1]. This implies that the women are either stagnating at lower positions or dropping out of the organizations. In India, women are apportioned with childcare and domestic responsibilities irrespective of their work status. Women give more precedence to family than career [5]. Hence, work-life balance (WLB), i.e., managing the job, family, and social responsibilities along with self-care, is a major issue afflicting working women [6]–[8].

Among the methods prevalent in assessing the WLB, statistical methods play an important role. The prediction can be made by two types of models. The first category includes statistical models, such as regression and structural equation modeling. These have linear structures and hence may have low prediction accuracy. The second group includes machine learning (ML) models, which are helpful to model nonlinear systems and often have higher prediction accuracy [10]–[12]. ML techniques are helpful to generalize new data effectively [13]. In many cases, ML techniques outperform multiple regression [12]. ML shows a strong capacity for abstraction, which depends on the choice of network topology and training methodology used. Their flexibility and robust nonlinear modeling lead to superior performance in prediction. ML is highly efficient in dealing with missing data. Nonlinear patterns are effectively captured compared to any other statistical technique. There are many advantages of ML compared to multiple regression such as no requirement of the predefined data model, learning from the data patterns, robustness against outliers', etc. However, multiple regression has an advantage when the phenomenon is to be explained rather than only predictions or classifications to be made

ML has been used increasingly in business research for decisionmaking. Organizational sciences are

benefitting through a prudent application of ML techniques. A study by Somers [10] on work attitudes that affect job performance involved a comparative study on the prediction capability of ML models and 8049regression. It stated how ML models can be a good alternative to traditionally used techniques such as regression. Shankar et al. [11] used ML classifiers, such as support vector machines (SVM), logistic regression, and decision trees, for the classification of employees based on the prediction of attrition intention. A similar study done by Yadav et al. [14] using SVM and random forest was done to predict employee attrition. A recent study taken up by Khera and Divya [15] uses ML to find employee attrition in the IT industry. One of the studies in Poland predicted WLB of employees using an ML approach [16]. It was one of the first studies on WLB where a huge data set of 800 employees was analyzed using ML. The ML cross-validation method was used by Yedida et al. [17] to predict whether an employee faces attrition.

Input Design:

In an information system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.

Therefore, the quality of system input determines the quality of system output. Well-designed input forms and screens have following properties −

- It should serve specific purpose effectively such as storing, recording, and retrieving the information.
- It ensures proper completion with accuracy.
- It should be easy to fill and straightforward.
- It should focus on user's attention, consistency, and simplicity.
- All these objectives are obtained using the knowledge of basic design principles regarding −
    - What are the inputs needed for the system?
    - How end users respond to different elements of forms and screens.

Objectives for Input Design:

The objectives of input design are −

- To design data entry and input procedures
- To reduce input volume

- To design source documents for data capture or devise other data capture methods
- To design input data records, data entry screens, user interface screens, etc.
- To use validation checks and develop effective input controls.

Output Design:

The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts.

Objectives of Output Design:

The objectives of input design are:

- To develop output design that serves the intended purpose and eliminates the production of unwanted output.
- To develop the output design that meets the end user's requirements.
- To deliver the appropriate quantity of output.
- To form the output in appropriate format and direct it to the right person.
- To make the output available on time for making good decisions.

LITERATURE SURVEY

[1]. Raghuram, P., Herman, C., Ruiz-Ben, E., and Sondhi, G. (2017), Women and IT Scorecard – India. Milton Keynes, U.K.: The Open University

In FY2018, India's IT-BPM industry size stood at USD 167 billion. The number of direct employees working in this industry stood at 3.97 million – India's largest private sector employer. The IT-BPM industry in India has always prided itself as being gender neutral – the focus has always been on the skills that you possess. Today, this industry has over 34% women workforce (1.3 to 1.4 million). However, while women form a substantial portion of the workforce, their presence in the boardroom is still far from adequate. As this report indicates, while there are large numbers of women entering the industry, very few make it to the very top. To improve this ratio, the Government of India has issued a directive that all listed companies must have at least one women director on their boards. While this is a significant move, a lot more needs to be done to enable women in leadership and executive roles.

Summary: It aims to understand and analyse the profile of women employees in India's technology sector covering demographic insights, career path, policies and practices, international mobility trends and also offers recommendations to individual firms and the industry as a whole.

[2]. Shanker, D. (2008), "Gender relations in IT companies: An Indian experience," Gender, Technology and Development, 12(2), 185–207

The emergence of Information Technology sector in mid-1990s has unveiled a potential employment opportunity for women in this organized sector congenially befitting their job environment and offering, in principle, least gender discrimination. This paper attempts to analyze the issues of opportunities and constraints the women employees face in the Information Technology sector in India. IT sector, through its employment, contributes substantially to women empowerment. Its employment potentiality provides inspiration to female students to take up technical and professional courses with an eye to the job market. Most reviews reveal that, notwithstanding overall satisfactory gender neutral pursuit by this sector, an optimal level of gender inclusivity is still to be achieved, especially to the senior level. Moreover, this sector requires to be extra-careful in doing away with the prevailing maladies such as 'Feminization', 'Glass ceiling' etc. The theoretical aspect of individualization in the workplace is palpable but at the societal level, patriarchal strategies dominate on the Indian psyche. The reflection of this paper is arrived at, on the basis of, the inputs drawn from different literatures of secondary sources.'

Summary: Achieving gender equality requires women's empowerment to ensure that decision-making at private and public levels, and access to resources are no longer weighted in men's favour, so that both women and men can fully participate as equal partners in productive and reproductive life.

[3]. Varma, R. (2002), "Women in information technology: A case study of undergraduate students in a minority-serving institution," Bulletin of Science,Technology & Society, 22(4), 274–282

The issue of underrepresentation of women in information technology (IT) is of national interest due to the rapid growth of IT in recent years, the impact of IT on growth and productivity, the shortage of IT workers, and the gender equity in IT. Scholarly research has pointed its finger at bias in early socialization,

math anxiety, masculinity of computers, the scarcity of role models, and women's preference for relational work. A study of students majoring in computer science and computer engineering in a minority-serving university found some additional factors such as tension between demands imposed by IT curricula and students'family and work responsibilities.

Summary : Minority students—those who do not belong to a region's or nation's majority racial or ethnic group—may be subject to discrimination, whether sanctioned or passive, that can affect their educational achievement.

4. Upadhya, C., and Vasavi, A. R. (2006), "Work, culture and sociality in the Indian information technology (IT) industry: A sociological study," National Institute of Advanced Studies, Bangalore, India

This report summarises the main conclusions of a sociological research of the Indian IT workforce that was conducted over more than two years, between November 2003 and March 2006, in Bangalore and three European nations. The goal of the research project, "Indian IT Professionals in India and Europe: Work, Culture, and Transnationalism," was to document the social and cultural changes brought about by the explosive growth of the IT and ITES sectors in India and, in doing so, to provide some insight into broader processes of globalisation. The study concentrated on the development of the IT workforce, the new forms of employment, organisation, and management, as well as the new work cultures that have emerged.

Summary: The Indian IT industry is known for its strong work ethic and focus on elivering quality results. This is reflected in the long hours and high level of dedication shown by employees. The culture is also collaborative, with teams often working closely together to achieve common goals.

EXISTING SYSTEM

In the existing system, we have implemented the support vector machine to visualize the r2 score. To overcome all this, we use machine learning packages available in the scikit-learn library.

Disadvantages:

- High complexity.
- Time consuming.

PROPOSED SYSTEM

Proposed several machine learning models to classify classify the score of the women's work

life balance who work in IT tech-industries by using machine learning. Therefore, we propose a Random Forest, Logistic Regression and Decision Tree machine Classifier to predict the work life balance women's.

Advantages:

- Highest accuracy
- Reduces time complexity.
- Easy to use

MODULES:

1. **User**:

1.1 View Home page:

Here user view the home page of the women's work life balance application.

1.2 View about page:

In the about page, users can learn more about the work life balance of women's.

1.3 View load page:

In the load_data page , the user will load the dataset for modelling.

1.4 Input Model:

The user must provide input values for the certain fields in order to get results.

1.5 View Results:

User view's the generated results from the model.

1.6 View score:

Here user have ability to view the accuracy score in %

2. **System**

2.1 Working on dataset:

System checks for data whether it is available or not and load the data in csv files.

2.2 Pre-processing:

Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

2.3 Training the data:

After pre-processing the data will split into two parts as train and test data before training with the given algorithms.

2.4 Model Building

To create a model that predicts the personality with better accuracy, this module will help user.

2.5 Generated Score:

2.6 Here user view the score in %

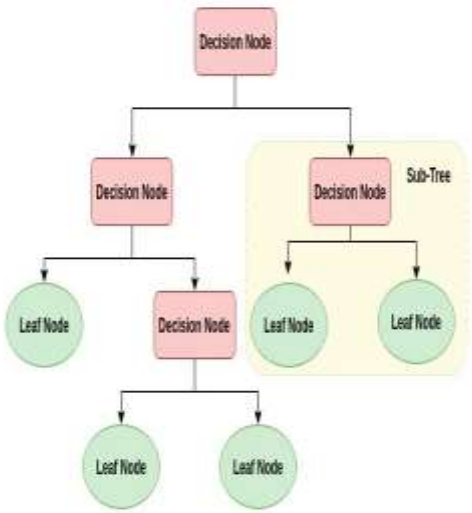
2.7 Generate Results:

We train the machine learning algorithm and predict the Students adaptability level.

METHODOLOGY AND ALGORITHMS:

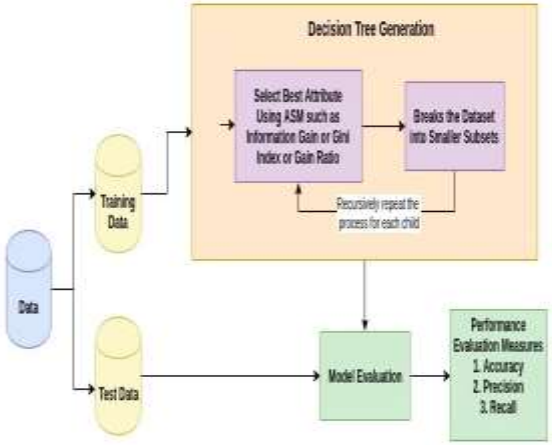
1. DECISION TREE:

Decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.

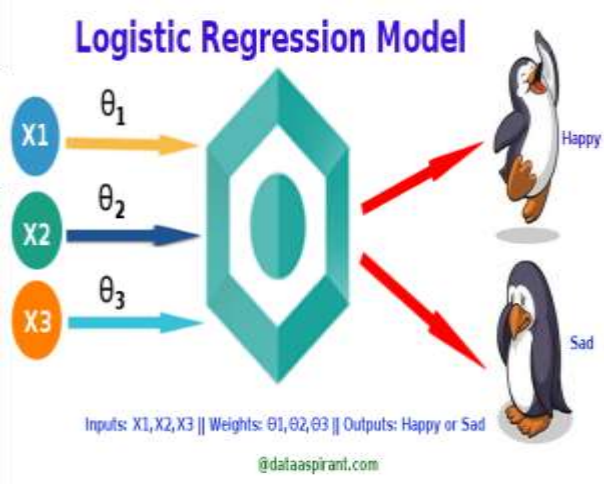


The basic idea behind any decision tree algorithm is as follows:

1. Select the best attribute using Attribute Selection Measures (ASM) to split the records.
2. Make that attribute a decision node and breaks the dataset into smaller subsets.
3. Starts tree building by repeating this process recursively for each child until one of the conditions will match:
    - All the tuples belong to the same attribute value.
    - There are no more remaining attributes.
    - There are no more instances.



2. Logistic Regression:



Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical. For example,

- To predict whether an email is spam (1) or (0)
- Whether the tumor is malignant (1) or not (0)
- Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.
- From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

The three types of logistic regression

1. **Binary logistic regression** - When we have two possible outcomes, like our original example of whether a person is likely to be infected with COVID-19 or not.
2. **Multinomial logistic regression** - When we have multiple outcomes, say if we build out our original example to predict whether someone may have the flu, an allergy, a cold, or COVID-19.
3. **Ordinal logistic regression** - When the outcome is ordered, like if we build out our original example to also help determine the severity of a

COVID-19 infection, sorting it into mild, moderate, and severe cases.

3. Random Forest Classifier:

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like Scikit-learn).

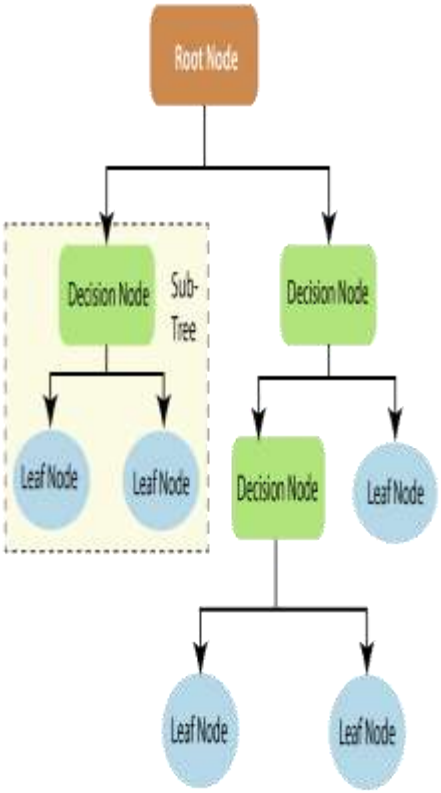
Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. An overview of decision trees will help us understand how random forest algorithms work.

A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further.

The nodes in the decision tree represent attributes that are used for predicting the outcome. Decision nodes provide a link to the leaves. The following diagram shows the three types of nodes in a decision tree.



The information theory can provide more information on how decision trees work. Entropy and information gain are the building blocks of decision trees. An overview of these fundamental concepts will improve our understanding of how decision trees are built.

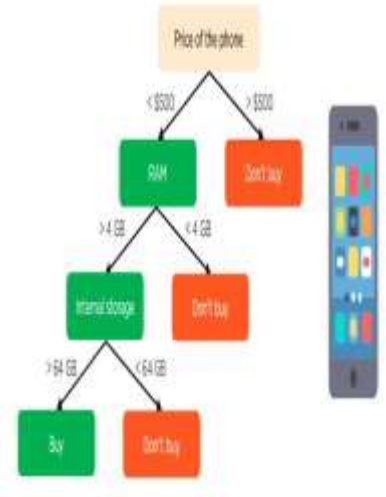
Entropy is a metric for calculating uncertainty. Information gain is a measure of how uncertainty in the target variable is reduced, given a set of independent variables.

The information gain concept involves using independent variables (features) to gain information about a target variable (class). The entropy of the target variable (Y) and the conditional entropy of Y (given X) are used to estimate the information gain. In this case, the conditional entropy is subtracted from the entropy of Y.

Information gain is used in the training of decision trees. It helps in reducing uncertainty in these trees. A high information gain means that a high degree of uncertainty (information entropy) has been removed. Entropy and information gain are important in splitting branches, which is an important activity in the construction of decision trees.

Let's take a simple example of how a decision tree works. Suppose we want to predict if a customer will purchase a mobile phone or not. The features of the phone form the basis of his decision. This analysis can be presented in a decision tree diagram.

The root node and decision nodes of the decision represent the features of the phone mentioned above. The leaf node represents the final output, either *buying* or *not buying*. The main features that determine the choice include the price, internal storage, and Random Access Memory (RAM). The decision tree will appear as follows.



Applying decision trees in random forest

The main difference between the decision tree algorithm and the random forest algorithm is that establishing root nodes and segregating nodes is done randomly in the latter. The random forest employs the bagging method to generate the required prediction.

Bagging involves using different samples of data (training data) rather than just one sample. A training dataset comprises observations and features that are used for making predictions. The decision trees produce different outputs, depending on the training data fed to the random forest algorithm. These outputs will be ranked, and the highest will be selected as the final output.

Our first example can still be used to explain how random forests work. Instead of having a single decision tree, the random forest will have many decision trees. Let's assume we have only four decision trees. In this case, the training data comprising the phone's observations and features will be divided into four root nodes.

The root nodes could represent four features that could influence the customer's choice (price, internal storage, camera, and RAM). The random forest will split the nodes by selecting features randomly. The final prediction will be selected based on the outcome of the four trees.

The outcome chosen by most decision trees will be the final choice. If three trees predict *buying*, and one tree predicts *not buying*, then the final prediction will be *buying*. In this case, it's predicted that the customer will buy the phone.

Support Vector Machines

Support Vector Machine is widely regarded to SVM, which is nothing but acronym of it. SVM is one among the mostly known supervised kind of learning methodologies which is much utilized for the sake of sorting out the issues related to regression/ classification. Nonetheless, when it comes to ML, SVM is widely found deployed for sorting out any classification issues from perspectives of different applications [13]. The main aim of these approaches is to discover a hyper place in the space of N-dimension that is able to

categorize the particulars of data in an unique way [14].

From general sense, the SVMs are of two types. The two types are as follows: Non-Linear SVMs and Linear SVMs.

SVM Kernel

For easing the data manipulation operations in SVMs, SVM kernels are utilized. It is nothing but a function which considers the space of the lower dimensional input and converts it into a maximum space of dimensional input. In other words, it could be said that it transforms the non-segregable issue into segregable issue [15]. This SVM kernel is found much beneficial in the non-linear segregation issues. The main purpose of using this SVM kernel is to carry out a few sophisticated/ difficult data conversions and subsequently discover the operation by which the data segregation is done depending upon the defined outcomes or labels [15].

Applications of support vector machine

According to [16], the application of SVMs are discussed in the below pointers.

- **Hypertext as well as Text organization -** As the classification methodology, it is deployed to either discover significant data or one could tell the needed data to organize the text contents.
- **Face examination -** It is deployed to identify the face of the human beings in accordance with the model as well as the devised classifier.
- **Bio-informatic context –** SVMs are utilized for the context of medical field and also in laboratories and heath care facilities. For instance, SVMs are found utilized for the locating and classifying purposes when the sequences of amino acids are known.
- **Grouping -** SVMs are found deployed for several grouping needs. For instance, a decision can be arrived just by comparing any data to facilitate grouping to certain or needed class.
- **Recall of Hand writing -** SVMs are utilized for identifying the hand writing of any human beings by comparing the input and samples.

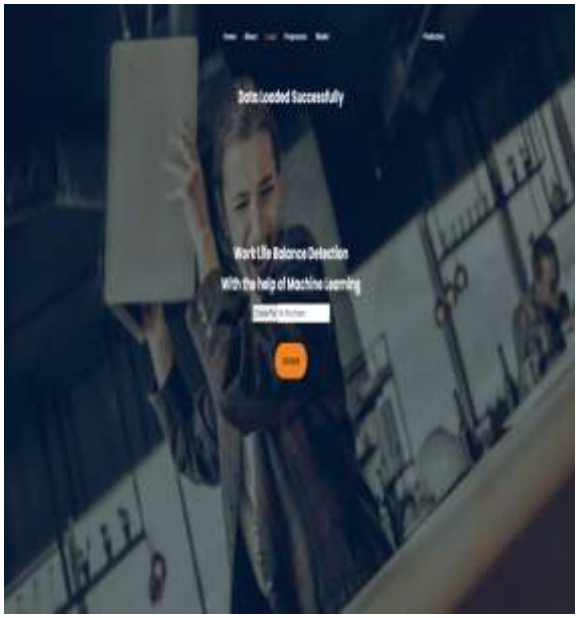
1.

SCREEN SHOT

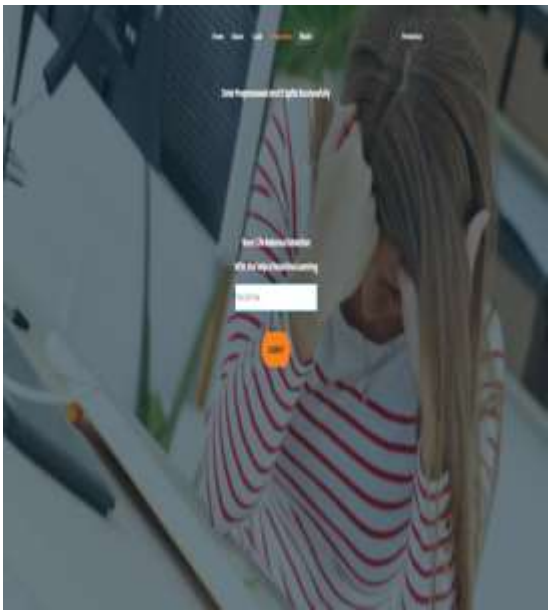
Home Page

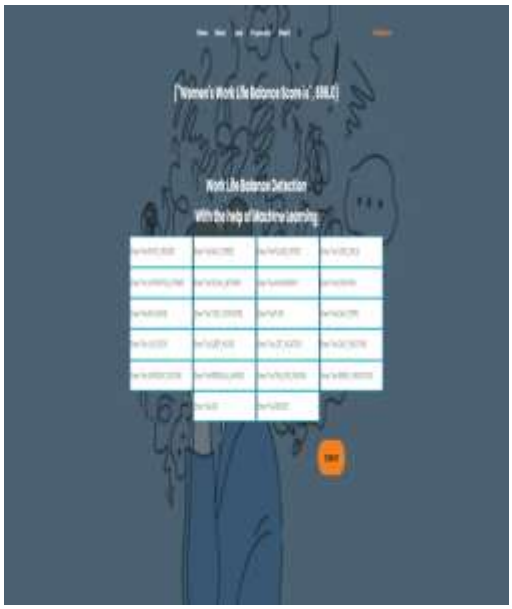
About Page:

Load



Preprocess





Model Selection

Prediction

CONCLUSION:

By using AI intelligence we have implemented machine learning techniques through detection of the concept of work-life balance for women often includes additional challenges and responsibilities, such as caregiver duties and gender-based wage disparities. Despite advancements in gender equality, women still tend to bear a disproportionate burden of household and caregiving tasks, making it difficult to achieve a healthy balance between work and personal life. To address these challenges, it is important for employers to offer flexible work arrangements, support

equal pay and opportunities, and provide resources for work-life balance. It is also crucial for women to prioritize self-care, seek support from family and friends, and advocate for their needs at work. Improving work-life balance for women can lead to improved health, increased job satisfaction, and a more equitable society.

REFERENCES:

[1] Raghuram, P., Herman, C., Ruiz-Ben, E., and Sondhi, G. (2017), Women and IT Scorecard – India. Milton Keynes, U.K.: The Open University

[2] Shanker, D. (2008), "Gender relations in IT companies: An Indian experience," Gender, Technology and Development, 12(2), 185–207.

[3] Varma, R. (2002), "Women in information technology: A case study of undergraduate students in a minority-serving institution," Bulletin of Science, Technology & Society, 22(4), 274–282.

[4] Upadhya, C., and Vasavi, A. R. (2006), "Work, culture and sociality in the Indian information technology (IT) industry: A sociological study," National Institute of Advanced Studies, Bangalore, India.

[5] Bhatnagar, D., and Rajadhyaksha, U. (2001), "Attitudes towards work and family roles and their implications for career growth of women: A report from India," Sex Roles, 45(7/8), 549–565.

[6] Agarwal, S., Jindal, A., Garg, P., and Rastogi, R. (2017), "The influence of quality of work life of trust: Empirical insights from an SEM application," International Journal of Indian Culture and Business Management, 15(4), 506–525.

[7] George, J., and Wallio, S. (2017), "Organizational justice and millennial turnover in public accounting," Employee Relations, 39(1), 112–126.

[8] Singh, T., Nandan, S., and Chawla, G. (2015), "HR service dimensions of quality of work life factors: ITenabled services perspectives in India," International Journal of Indian Culture and Business Management, 10(4), 460–475.

[9] Ahuja, M. K. (2002), "Women in the information technology profession: A literature review, synthesis and research agenda," European Journal of Information Systems, 11(1), 20–34.

[10] Somers, M. J. (2001), "Ethical codes of conduct and organizational context: A study of

the relationship between codes of conduct, employee behavior and organizational values," Journal of Business Ethics, 30(2), 185–195.