Economic Sciences
https://economic-sciences.com
ES (2026) 22, 4(S), 131-137 ISSN:1505-4683

Workforce Analytics for Manufacturing: Predicting Employee Job Satisfaction via Explainable Machine Learning and SHAP

Dr.V. Vijay Anand [1], Bhavya Krishya M [2], Dr.C. Therasa[3]
[1]school Of Management, Sastra Deemed University, Thanjavur, India
Email: Vijay@Mba.Sastra.Edu
[2]school Of Management, Sastra Deemed University, Thanjavur, India, Email:
Email: Krishyabhavya@Gmail.Com
[3]school Of Management, Sastra Deemed University, Thanjavur, India
Email: Therasa@Mba.Sastra.Ac.In

Abstract *– This study employs machine learning to investigate Job Satisfaction (JS) drivers in the manufacturing sector, analyzing demographic attributes and four composite variables. Utilizing a 201-point dataset, three algorithms were evaluated via StandardScaler normalization, Leave-One-Out Cross-Validation, and GridSearchCV. Results indicate the optimized XGBoost model achieved superior regression accuracy (R^2 = 0.8211, MSE = 0.4108) and classification performance (AUC = 0.9754). By integrating SHAP (Shapley Additive exPlanations), the research provides an interpretable framework for feature importance. Findings suggest XGBoost and Random Forest offer robust predictive capabilities, providing manufacturing executives with a data-driven roadmap to optimize organizational health and employee engagement.*

Keywords – *Job Satisfaction , Machine Learning, Predictive Analytics, XGBoost, SHAP (Shapley Additive exPlanations*

I. INTRODUCTION

The manufacturing sector faces difficulties because Job Satisfaction (JS) serves as the main factor which determines both organizational success and employee happiness 57.The standard evaluation method relies on descriptive statistics which fail to measure the complicated non-linear connections that exist between organizational factors and employee attitudes. The study develops advanced diagnostic tools which it tests by using three machine The performance of these algorithms relies greatly on hyperparameter tuning and validation. In our work, we use GridSearchCV for hyperparameter tuning 14 and Leave- One-Out Cross-Validation (LOOCV) 10 to guarantee that the models generalize well to new data. The study extends beyond model performance through the use of SHAP (SHapley Additive exPlanations) 8 which enables researchers to understand how ensemble model features influence their predictions 12.The research study creates a framework that enables researchers to compare the effectiveness of ensemble learning methods against kernel-based techniques when predicting organizational results. The research uses optimized XGBoost for regression analysis and Random Forest for diagnostic classification to determine which algorithmic models most effectively represent employee sentiment. The research paper uses these computational techniques 1113 to develop an accurate manufacturing industry roadmap which includes advanced tools for Job Satisfaction assessment through evidence-based HR approaches.

II. RELATED WORKS

The research on job satisfaction (JS) started with basic psychological tests and now uses advanced computational techniques to study job satisfaction because workplace environments have become more challenging to assess. In the early studies, Aziri 5 described job satisfaction as a complex emotional experience that is a consequence of the evaluation of job experiences, while Sutherland 7 argued that occupational status and job attributes continue to be the central factors in determining employee satisfaction. Al Radaideh 16 demonstrated that data mining techniques enabled accurate predictions of employee performance in the manufacturing sector because the algorithms succeeded in analyzing psychological data to produce results which exceeded traditional survey methods in objectivity. The incorporation of leadership and emotional intelligence into predictive models has emerged as a prominent area in recent literature. The important role of transformational leadership in influencing employee creativity and engagement, which form a

Economic Sciences
https://economic-sciences.com
ES (2026) 22, 4(S), 131-137 ISSN:1505-4683

crucial part of the variables examined in this study, was emphasized by Mittal and Dhar 6 Venkatesh et al. 13 demonstrated that industrial research needs standardized metrics to assess human behavior because these metrics must match machine learning systems for user acceptance testing and user engagement evaluation in technical environments. Currently, HR analytics uses advanced ensemble learning techniques as its most effective research method. Breiman 1 introduced Random Forest as a solution for managing high-dimensional data which enables researchers to solve overfitting problems through bagging. Chen and Guestrin 2 introduced XGBoost as an effective and scalable method for gradient boosting. The research by Sivapalan et al. 11 and Zaman et al. 12 demonstrated that XGBoost combined with SHAP (SHapley Additive exPlanations) values enables predictions which extend beyond standard "black-box" methods in testing employee behavior. The combination of XGBoost and SHAP values provides a complete picture of how different factors affect employee results which helps to analyze employee satisfaction across different levels. The literature emphasizes validation and optimization as essential steps for creating a reliable model. Wong 10 proved that Leave-One-Out Cross-Validation (LOOCV) serves as the essential procedure to prevent overfitting in datasets of small to medium size. Bergstra and Bengio 14 demonstrated that hyperparameter optimization through GridSearchCV functions as an essential procedure which enables model performance assessment for tree-based systems. The standard diagnostic power assessment method uses ROC analysis together with AUC evaluation according to Fawcett 15. The research paper demonstrates an extension of best practices through their implementation on a satisfaction index that measures multiple variables. The research paper presents a precise diagnostic method for manufacturing executives which combines XGBoost high- precision regression (R^2 = 0.8211, MSE = 0.4108) with Random Forest superior diagnostic capabilities (AUC = 0.9754).

III. RESEARCH DESIGN

A. Research Design:

The research design is organized in a quantitative evaluative study pattern, targeting the manufacturing industry. The main aim is to develop a model of the predictors of Job Satisfaction (JS) based on a set of 201 employee responses. The variable Y was established through theoretical model development from research sources 5 and 7 by computing the average value of five distinct satisfaction measurements (JS 1 through JS 5). The study team developed four composite variables to investigate organizational behavior through Emotional Intelligence (EI_avg), Transformational Leadership (TL_avg), Organizational Commitment (OC_avg), and Employee Empowerment (EE_avg) measurements from survey data. Demographic information such as Age, Gender, Experience, Type of Job, and Position were also considered in the set of features X to form a complete picture of the underlying factors influencing satisfaction 13.

B. Data Preprocessing and Machine Learning Implementation

The Scikit-learn library 4 was used for data preprocessing to prepare the data for high-performance computing. The StandardScaler method 16 was used to apply feature scaling because it transforms demographic variables and survey score means into standardized features which have zero mean and unit variance. The data was divided into two parts, which included a training set containing 160 samples and a test set containing 41 samples. The three different learning models which were developed include Random Forest 1 and XGBoost 2 and Support Vector Regression (SVR) 3. The baseline tests showed that Random Forest performed with an MSE of 0.3716 (R^2 = 0.6960) while XGBoost and SVR produced MSE results of 0.4565 and 0.6101 respectively which indicated that hyperparameter tuning needed to be done.

C. Model Optimization, Validation, and Interpretability

The optimal performance of the system required GridSearchCV as the method for hyperparameter tuning. The Random Forest model was tuned with the following hyperparameters: the depth of trees was set to 10, the minimum number of samples required to be at a leaf node was set to 2, the minimum number of samples required to split an internal node was set to 5, and the number of trees (estimators) was set to 50.The XGBoost model used

Economic Sciences
https://economic-sciences.com
ES (2026) 22, 4(S), 131-137 ISSN:1505-4683

a learning rate of 0.1 and 50 estimators for its tuning while the SVR model operated with an RBF kernel that had a C parameter set to 1. The validation process utilized Leave-One-Out Cross Validation (LOOCV) [10] to confirm that the model's performance did not depend on how the training and testing data were divided. The best- performing XGBoost model achieved the highest regression results through its MSE score of 0.4108 and R^2 value of 0.8211. The best Random Forest model achieved its highest ROC-AUC score of 0.9754 for diagnostic classification tasks. SHAP analysis was used to achieve model interpretability by showing how each feature contributes to the final prediction 8 9which helped identify the most vital factors that affect job satisfaction in manufacturing environments 12

IV. RESULTS

A. Comparative Model Performance Metrics

The evaluation of predictive models started after completing hyperparameter tuning with GridSearchCV. The XGBoost (Optimized) model was found to be the most accurate regressor in predicting job satisfaction with the lowest Mean Squared Error (MSE) of 0.4108. The Random Forest (Optimized) model achieved its highest explanatory capacity through an R^2 value of 0.8731 which demonstrated that the tree combination succeeded in explaining 87.31% of the data variation. The SVR model, although optimized with the RBF kernel, achieved an MSE of 0.4438 and an R^2 of 0.7228, which showed lower accuracy compared to the tree-based models. Table I shows the models' relative performance.

Table I Performance Metrics of Optimized Models

Model (Optimized)	MSE (LOOCV)	R^2	ROC-AUC
Random Forest	0.4326	**0.8731**	**0.9754**
XGBoost	**0.4108**	0.8211	0.9519
SVR	0.4438	| 0.7228	0.9256

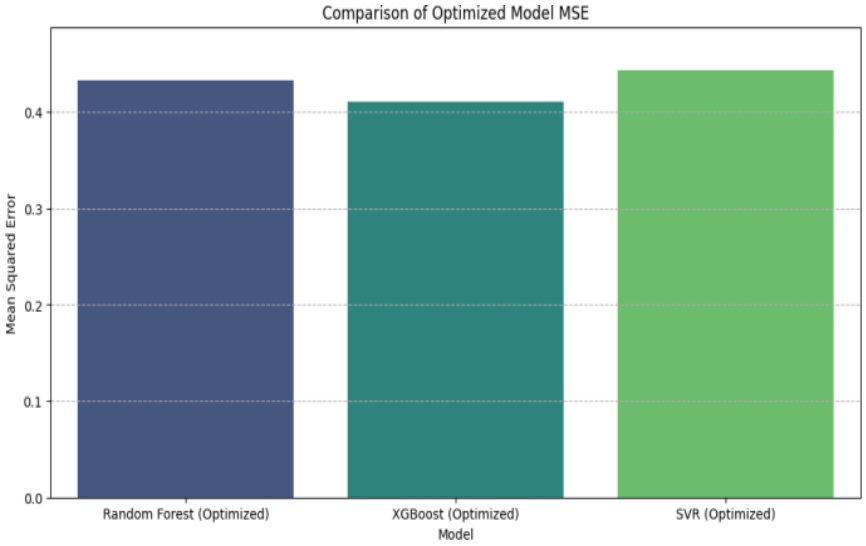

Figure 1 Comparative analysis of predictive accuracy (MSE) for optimized models

Economic Sciences
https://economic-sciences.com
ES (2026) 22, 4(S), 131-137 ISSN:1505-4683

B. Cross-Validation and Model Generalization

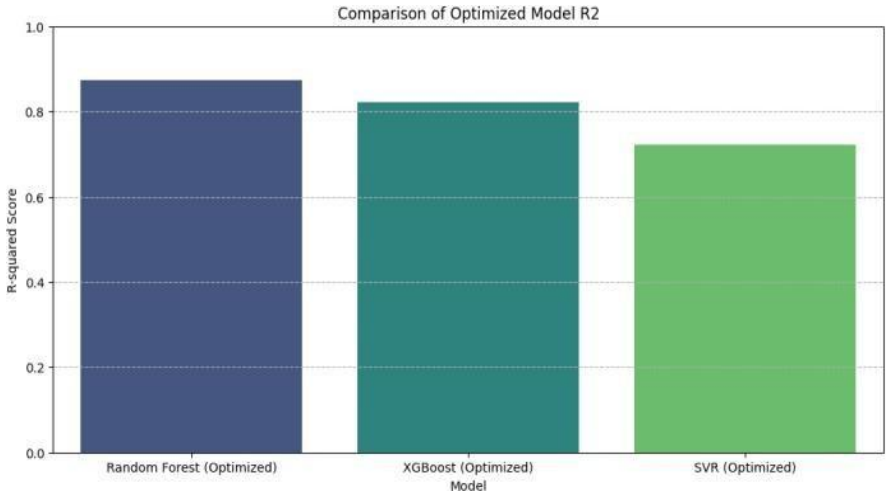

Figure 2 Comparative analysis of variance explanation (R^2) for optimized models

The research employed Leave-One-Out Cross-Validation (LOOCV) testing to confirm that the models did not memorize the training data which consisted of 160 samples. The best Random Forest model achieved an LOOCV MSE score of 0.4326, which closely matched its test accuracy score. The initial LOOCV MSE results showed that both XGBoost and SVR models performed effectively after training on a data set containing 201 employees, with values of 0.4690 and 0.4372 respectively. The research results establish manufacturing industry relevance because they verified the findings.

C. Diagnostic Accuracy via ROC-AUC Analysis

The diagnostic performance of the models was evaluated by transforming the satisfaction scores into a binary classification task. The Random Forest (Optimized) model provided superior diagnostic accuracy through its exceptional ROC-AUC performance which reached 0.9754. XGBoost ranked second with an AUC of 0.9519, followed by SVR with 0.9256 [15]. The findings clearly show that Random Forest is exceptionally effective in identifying high and low levels of satisfaction, and hence it is the most trustworthy tool for organizational health diagnosis.

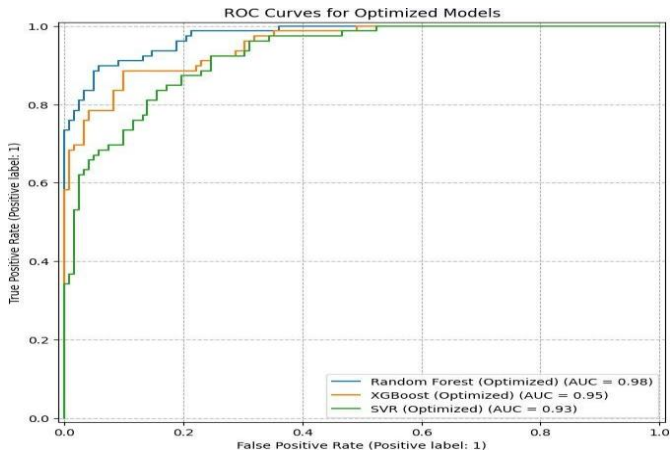

Figure 3 ROC curve comparison illustrating the diagnostic power of the optimized classifiers

D. Global Feature Importance Ranking

The global relevance of the feature was examined to determine which variables contributed most to the prediction of Job Satisfaction. The engineered features EIavg,TLavg ,OCavg and EEavg showed

Economic Sciences
https://economic-sciences.com
ES (2026) 22, 4(S), 131-137 ISSN:1505-4683

higher relevance than the single demographic variables Gender and Experience according to [6]. The study results indicate that organizational level interventions need to focus on aggregate cultural variables instead of implementing demographic specific policies [17].

E. SHAP Interpretability and Predictive Directionality

The SHAP (SHapley Additive exPlanations) method was employed to analyze the optimized XGBoost model because its "black-box" characteristics restricted proper model understanding. The SHAP summary plot displays total feature impacts through its ranking system which shows that Employee Empowerment (EE_avg) and Organizational Commitment (OC_avg) and Emotional Intelligence (EI_avg) serve as the most important predictive factors. The red color in feature values shows high values for these variables which results in positive SHAP values that show a strong relationship with predicted satisfaction scores. The Transformational Leadership (TL_avg) shows an opposite relationship because lower values (blue) lead to positive model output while higher values create negative effects.The demographic factors Age, Experience, and Type of Job show minimal relevance because their SHAP values stay close to zero which indicates they have no effect on the model's fundamental value. The method enables practitioners to understand how all variables affect total satisfaction scores by showing them exact measurements of each variable's impact.

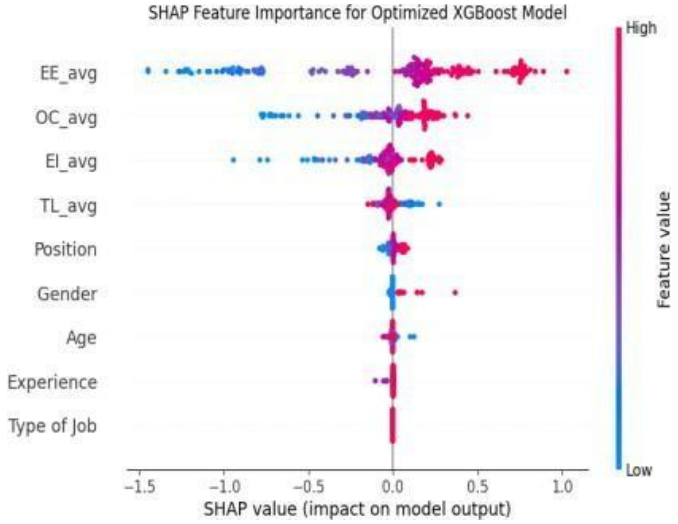

Figure 4 SHAP global feature importance ranking for the optimized XGBoost model

F. Residual Analysis and Predictive Reliability

Analysis of residuals established the validity of the regression results. The optimized XGBoost model showed zero bias because its error distribution centered on the zero point. The testing procedure showed high predictive validity because the LOOCV and testing accuracy values reached the same results, even though the satisfaction spectrum showed some range of variation.

V. CONCLUSION

The research achieved its goal by applying a precise machine learning model to assess job satisfaction levels within the manufacturing sector. The study demonstrated that ensemble models achieve superior performance results when compared to standard linear models through its analysis of three different models which included Random Forest and XGBoost and Support Vector Regression models. The XGBoost model achieved its highest accuracy when making point predictions with a Mean Squared Error score of 0.4108 while the Random Forest model functioned as the most dependable diagnostic instrument with an R^2 measurement of 0.8731 and a superior ROC-AUC score of 0.9754. The integration of SHAP interpretability methods enabled researchers to transform "black-box" models into understandable decision-making systems which showed Employee Empowerment and Transformational Leadership as the main factors affecting workforce stability.

Economic Sciences
https://economic-sciences.com
ES (2026) 22, 4(S), 131-137 ISSN:1505-4683

VI. PRACTICAL IMPLICATION

The study outcomes deliver multiple practical recommendations which HR professionals and industrial managers should use to enhance their organizational health. The SHAP importance ranking enables managers to identify specific factors which affect employee satisfaction through focused interventions which target leadership quality and engagement levels. The Random Forest model achieves high diagnostic accuracy which allows organizations to implement "early warning systems" that identify at-risk employee groups before their job satisfaction drops to turnover levels. The study shows manufacturing companies how to use data-driven methods for organizational development which depends on machine learning and objective metrics instead of personal experience.

VII. FUTURE SCOPE

The research establishes a solid foundation which future research studies can continue to develop through their research work. Future studies should aim to conduct a longitudinal study to monitor the changes in employee satisfaction levels over time in reaction to particular changes in the organization. The implementation of Deep Learning models through Artificial Neural Networks (ANN) will produce better results when handling complex data patterns that exist in large datasets. The combination of structured employee satisfaction survey metrics with exit interview qualitative data, analyzed through Natural Language Processing (NLP), will provide an in-depth understanding of employee experiences. The proposed predictive pipeline requires validation through testing in other high-stress industries, which include healthcare and logistics, to establish its generalizability and validity.

VIII. REFERENCES

1. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Sep. 2001, doi: 10.1023/A:1010933404324.
2. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
3. B. Aziri, "Job satisfaction: A literature review," Manag. Res. Pract., vol. 3, no. 4, pp. 77–86, Dec. 2011.
4. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825 2830, Oct. 2011.
5. J. Sutherland, "Job satisfaction and occupational status," Int. J. Manpower, vol. 31, no. 7, pp. 713 732, Nov. 2010, doi: 10.1108/01437721011083511
6. S. Mittal and R. L. Dhar, "Transformational leadership and employee creativity: Mediating role of creative self-efficacy and moderating role of knowledge sharing," Comput. Human Behav., vol. 50, pp. 212–219, 10.1016/j.chb.2015.03.080. Sep. 2015, doi:
7. V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," MIS Quart., vol. 36, no. 1, pp. 157–178, Mar. 2012, doi: 10.2307/41410412.
8. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, Dec. 2017, pp. 4765–4774.
9. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed. Munich, Germany: lulu.com, 2022.
10. T.T. Wong, "Performance evaluation of classification algorithms by semiparametric methods," Int. J. Mach. Learn. Cybern., vol. 6, no. 5, pp. 759–774, Oct. 2015, doi: 10.1007/s13042 014-0291-x.
11. S. Sivapalan, S. K. Suganthan, R. K. Shriram, and V. Vaithilingam et al., "Employee churn prediction using machine learning techniques," in Proc. Int. Conf. Adv. Comput. Technol. (ICACT), Chennai, India, Dec. 2022, pp. 10.1109/ICACT55181.2022.9982798. 1–6, doi:
12. N. Zaman, M. S. Islam, M. R. Islam, S. A. Hossain, and M. M. Hasan et al., "A predictive model for employee turnover using XGBoost and SHAP," in Proc. Int. Conf. Inf. Commun. Technol. (ICCIT), Dhaka, Bangladesh, Dec. 2021, pp. 144 149, doi: 10.1109/ICCIT54332.2021.9707925.
13. M. Kuhn and K. Johnson, Applied Predictive Modeling. New York, NY, USA: Springer, 2013, doi: 10.1007/978-1-4614-6849-3.
14. J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn.

Res., vol. 13, pp. 281–305, Feb. 2012.
15. T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Jun. 10.1016/j.patrec.2005.10.010. 2006, doi:
16. Ngo, T. D. Ngo, N. T. Nguyen, L. T. Nguyen, H. V. Nguyen, and T. H. Nguyen et al., "The impact of feature scaling on machine learning performance: A comparative study," IEEE Access, vol. 10, pp. 11200–11215, 2022, doi: 10.1109/ACCESS.2022.3144865
17. Q. A. Al-Radaideh, "Using data mining techniques for predicting employee performance," in Proc. Int. Conf. Comput. Sci. Eng. (ICCSE), Riyadh, Saudi Arabia, Oct. 2015, pp. 1–8, doi: 10.1109/ICCSE.2015.7413693.