e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

THE CORPORATE TRAINING ROI AND PERFORMANCE PERDICTION ENGINE

Bharath M*1, Dheekachanadevi S*2

*1PG Final Year Student, Department Of Computer Science, Rathinam College Of Arts And Science, Coimbatore, Tamil Nadu, India.

*2Technical Trainer, Department Of Computer Science, Rathinam College Of Arts And Science, Coimbatore, Tamil Nadu, India.

ABSTRACT

The Corporate Training ROI and Performance Prediction Engine is an intelligent, data-driven platform designed to assess and enhance the effectiveness of corporate training programs while predicting employee performance outcomes. By leveraging advanced analytics and machine learning, the system empowers HR teams, training coordinators, and managers to make informed decisions that drive continuous learning and organizational growth. The engine evaluates training impact through pre- and post-training performance comparisons, employee feedback sentiment analysis, skill development tracking, and engagement metrics. It features an interactive visualization dashboard that displays key insights and trends, helping stakeholders pinpoint successful modules and areas needing improvement. Furthermore, its built-in prediction engine employs regression and clustering models to forecast employee development trajectories and identify performance patterns across teams.

Keywords: Employee Training, Performance Prediction, Training Effectiveness, Machine Learning, Skill Development, HR Analytics, Data Visualization, Feedback Analysis, Corporate Learning, Streamlit, Python, NLP, ROI of Training

I. INTRODUCTION

The Corporate Training ROI and Performance Prediction Engine is a data-driven analytics platform designed to assess the effectiveness of corporate training programs while measuring their return on investment (ROI). Organizations invest heavily in employee training, but without proper evaluation, it becomes challenging to determine its actual impact. This system leverages advanced technologies such as machine learning and natural language processing (NLP) to analyze training outcomes, providing valuable insights into employee skill development, engagement levels, and overall performance improvements. One of the key functionalities of this platform is performance metrics analysis, which compares pre- and post-training performance data, allowing organizations to measure learning effectiveness objectively. Additionally, sentiment analysis of employee feedback helps HR teams and managers understand employee perceptions of the training programs. Using NLP techniques, the system analyzes qualitative feedback to identify key areas of satisfaction and areas that need improvement. Skill development tracking ensures that organizations can monitor employee progress over time, recognizing which training modules contribute to skill enhancement and which require modifications.

II. METHODOLOGY

The methodology began with collecting and cleaning data related to employee demographics, training sessions, and post-training performance metrics. Preprocessing steps included handling missing values, encoding categorical features, and scaling numerical data. Exploratory Data Analysis (EDA) was performed using visual tools like heatmaps, line charts, and bar graphs to uncover correlations and trends.

Data Collection

The data for this project was gathered from simulated corporate training logs designed to represent real-world scenarios across various departments and roles. These logs included essential employee details such as role, department, and years of experience, which are critical in determining the baseline and post-training performance.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

Data Preprocessing and Feature Engineering

To ensure the dataset was clean and consistent, missing values were handled appropriately—typically by imputation techniques—while duplicate records were removed to maintain data integrity. Categorical variables, such as job roles and training types, were encoded using One-Hot Encoding for non-ordinal categories and Label Encoding for ordinal data like experience levels.

Exploratory Data Analysis (EDA)

A thorough exploratory data analysis was conducted to gain insights into the structure and trends within the dataset. Bar charts and pie charts were used to visualize the distribution of training sessions across departments and employee levels.

Model Training

To build an effective prediction engine, three different machine learning models were developed and evaluated. The Linear Regression model served as a baseline, offering simple but interpretable results. The Random Forest Regressor was then implemented to capture more complex, non-linear relationships and assess feature importance. Finally, the XGBoost Regressor was selected for its superior performance in handling large datasets and variable interactions, ultimately delivering the highest predictive accuracy.

Predictive Analysis

The predictive analysis module allows users to input training parameters—such as cost, duration, and employee experience—to receive real-time forecasts of expected ROI and performance improvements. The model uses the same preprocessing and scaling pipeline as during training to ensure consistent predictions.

Risk Analysis

Risk analysis was performed to understand which variables posed the highest risk of ineffective training investment. Correlation analysis highlighted key cost drivers, while feature importance metrics (e.g., from Random Forest and SHAP values) indicated which inputs had the most influence on ROI outcomes.

Final Reporting

To ensure clarity and transparency in decision-making, the system generates automated PDF reports summarizing the training evaluation. These reports include sections such as a project overview, ROI forecasts, risk factors, and model-driven insights.

III. MODELING AND ANALYSIS

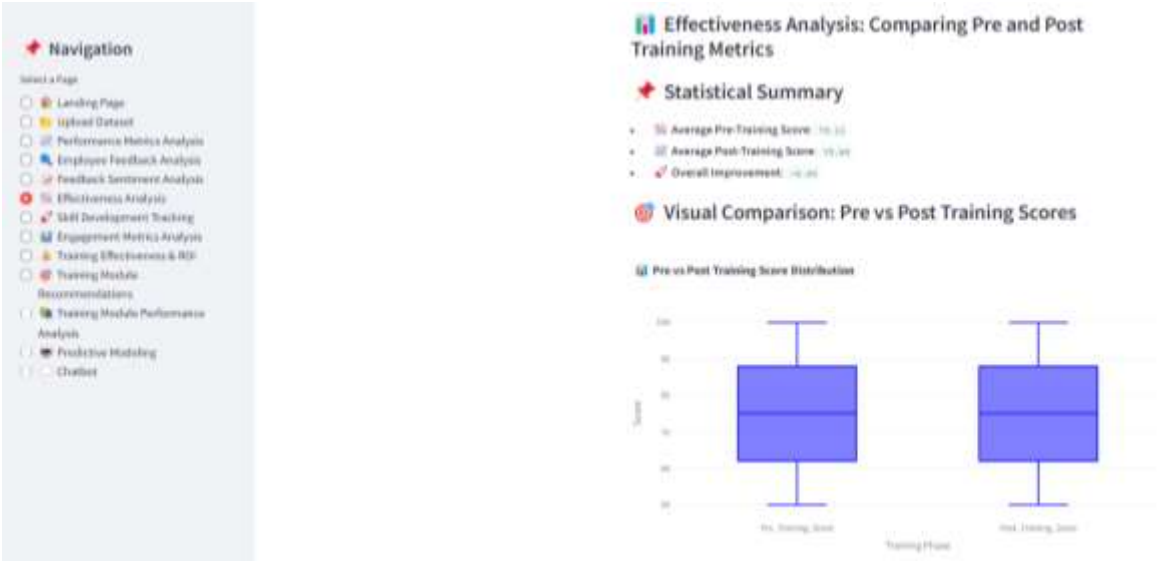

Figure 1: Effectiveness Analysis

This dashboard section analyzes the effectiveness of training by comparing pre- and post-training scores. The average pre-training score is 75.11, while the post-training score is slightly higher at 75.99. However, the overall improvement is minimal at -0.88, indicating a slight drop. A box plot visually compares score

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

distributions, showing a similar range for both phases. This suggests that the training program had little to no impact on performance.

IV. ARCHITECTURE DIAGRAM

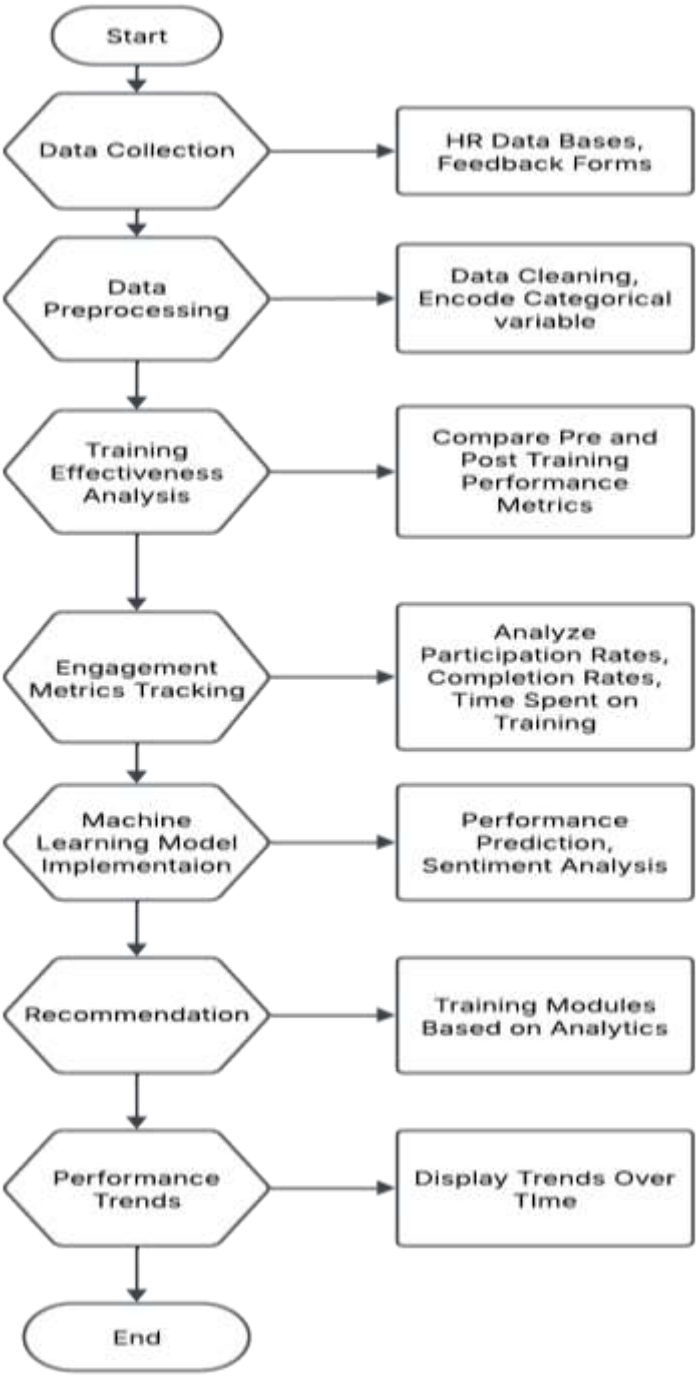

Figure 2: Architecture Diagram

V. RESULT AND DISCUSSION

This dashboard section shows predictive modeling to evaluate training effectiveness using Linear Regression. The model achieved high accuracy at **98.6%**, indicating strong predictive power. The **Mean Squared Error** is **214.32**, reflecting the average squared difference between actual and predicted values. The scatter plot compares actual vs. predicted scores, showing a close alignment. This suggests the model performs well in estimating post-training performance.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

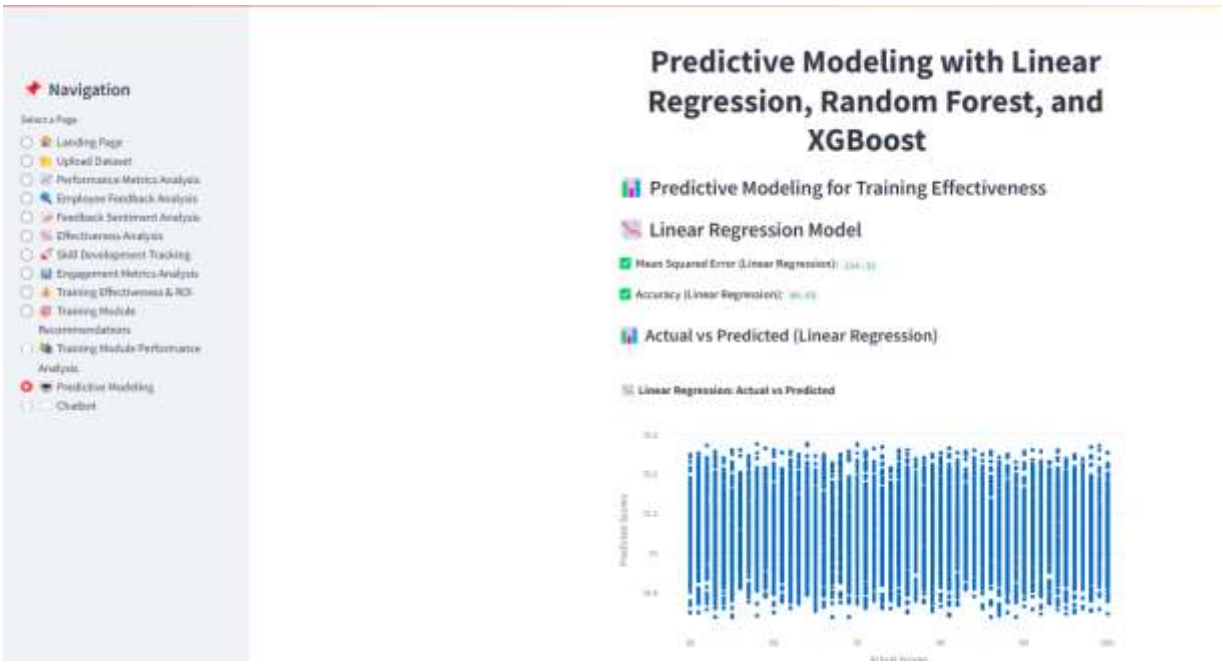

Figure 3: Predictive Analysis

VI. CONCLUSION

The Corporate Training ROI and Performance Prediction Engine is a powerful tool for organizations seeking to maximize the value of their training investments. With accurate forecasting, intelligent risk assessment, and an easy-to-use interface, it supports evidence-based planning and resource optimization. The integration of machine learning with business intelligence tools offers a comprehensive solution for measuring training success and aligning it with organizational goals. This project not only provides actionable insights but also sets the foundation for more advanced, AI-driven talent development systems in the future.

VII. REFERENCES

[1] Rothwell, W.J. & Kazanas, H.C., Mastering the instructional design process: A systematic approach. 2nd ed. San Francisco, CA: Jossey-Bass Publishers, 1998.

[2] Branson, R. K., "The interservice procedures for instructional systems development," Educational Technology, pp. 11-14, Mar 1978.

[3] Dick, W. & Carey, L., The systematic design of instruction, 4th ed. New York: Harper Collins Publishing, 1996.

[4] Molenda, M., In search of the elusive ADDIE model. Retrieved on June 9, 2010 from http://www.indiana.edu/~molpage/In%20Search%20of%20Elusive%20ADDIE.pdf#search=%22ADDIE%20Model%20%2Bhistory%22.

[5] Bramley, P. & Newby, A. C., "The evaluation of training part I: Clarifying the concept," Journal of European & Industrial Training, 8:6, pp. 10-16, 1984.

[6] Leigh, D., "A brief history of instructional design." Retrieved June 9, 2010 from http://www.pigncispi.com/articles/education/brief%20history.htm

[7] Becker, G.S., Human capital: A theoretical and empirical analysis with special reference to education, 3rd ed. Chicago, IL: University of Chicago Press, 1993.

[8] Phillips, J.J., Improving supervisors' effectiveness, San Francisco, CA: Jossey-Bass Publishers, 1985.

[9] Phillips, J.J., Recruiting, training and retaining new employees: Managing the transition from college to work, San Francisco, CA: Jossey-Bass Publishers, 1987.

[10] Boudreau, J.W. & Ramstad, P.M., Beyond HR: The new science of human capital, Boston, MA: Harvard Business School Press, 2007.

[11] Shadish, W.R., Cook, T.D., & Campbell, D.T., Experimental and Quasi-Experimental Designs for

Generalized Causal Inference. Boston: Houghton-Mifflin, 2002.

[12] Phillips, J.J., Return on investment in training and performance improvement programs: A step-bystep manual for calculating the financial return, Houston, TX: Gulf Publishing Company, 1997.

[13] Sugrue, B. & Riviera, J., State of the industry: ASTD's annual review of trends in workplace learning and performance, Alexandria, VA: American Society for Training and Development, 2005.

[14] Alliger, G. M., Tannenbaum, S. I., Bennett, W. Jr., Traver, H., & Shotland, A., "A metaanalysis of the relations among training criteria," Personnel Psychology, 50, pp. 341-358, 1997.

[15] Simonson, M.R., "Evaluating teaching and learning at a distance," New Directions for Teaching and Learning, 71, pp. 87-94, 1997