International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

EMPLOYEE TRAINING PREDICTION AND ANALYTICS

[1] V Soujanya, [2] A Harshini, [3] K Vaishnavi, [4] Nani Goud
[1]AssistantProfessor, [234]Students
Department of Data Science
Siddhartha Institute of Technology & Sciences, Narapally
vsoujanya.cse@siddhartha.org.in, 23TQ1A6709@siddhartha.co.in, 23TQ1A6725@siddhartha.co.in, 23TQ1A6761@siddhartha.co.in,

ABSTRACT

Employee training is an important part of organizational development because it helps employees improve their knowledge, skills, productivity, and job performance. Organizations generate training-related data such as employee demographics, department, job role, training programs, completion status, assessment scores, attendance, and performance ratings. Analyzing this data can provide useful insights into training effectiveness and employee development.

The **Employee Training Prediction and Analytics** system is designed to analyze employee training information and identify patterns related to training participation and performance. The system processes attributes such as employee ID, department, job role, training program, training duration, attendance, assessment scores, previous performance, and training completion status. Data preprocessing techniques are used to handle missing values, duplicate records, inconsistent formats, and invalid entries.

The system provides analytical insights such as training completion rate, average assessment score, employee participation, department-wise training performance, training effectiveness, and performance trends. Interactive dashboards can display these results through KPI cards, charts, graphs, tables, and filters. The system enables users to compare training outcomes across departments, roles, programs, and time periods.

The prediction module uses historical employee training data and Machine Learning techniques to predict outcomes such as training completion, assessment performance, or training effectiveness. Factors such as attendance, previous performance, training duration, employee experience, and department can be used as predictive features. The prediction results can help identify employees or training programs that may require additional support.

Overall, the proposed system combines employee training analytics, performance monitoring, visualization, and predictive modeling in a centralized platform. It can help organizations understand training patterns and evaluate learning outcomes using available historical data. Future enhancements can include personalized training recommendations, skill-gap analysis, real-time learning analytics, employee feedback analysis, and automated training alerts.

I. INTRODUCTION

Employee training plays an important role in improving workforce capabilities and supporting organizational growth. Organizations conduct different types of training programs to improve technical skills, communication, leadership, compliance, and job-specific knowledge. The increasing availability of employee and training data provides opportunities to evaluate the effectiveness of these programs.

Training outcomes can be influenced by several factors, including employee experience, attendance, training duration, previous performance, job role, department, learning format, and assessment results. Understanding these factors can help organizations identify patterns in employee participation and learning performance.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Traditional training management systems generally focus on registering employees, scheduling courses, tracking attendance, and recording completion status. Although these functions are useful for administrative purposes, they may provide limited analytical and predictive capabilities. HR teams may need separate tools to perform detailed comparisons and trend analysis.

Data analytics can help organizations calculate important training metrics such as completion rate, attendance rate, average assessment score, training participation, and department-wise performance. Interactive dashboards can present these metrics through charts, graphs, tables, and KPI cards, making training data easier to interpret.

Machine Learning can further enhance training analytics by predicting employee training outcomes and identifying patterns associated with successful course completion or assessment performance. Historical training data can be used to build predictive models. The proposed system combines training analytics and predictive techniques to provide a comprehensive framework for employee learning and development analysis.

II. LITERATURE SURVEY

Employee training analytics has become an important area of Human Resource Analytics because organizations collect increasing amounts of employee learning and development data. Training datasets may contain information about employee roles, departments, attendance, training programs, assessment scores, experience, and performance. These attributes can be analyzed to understand employee learning patterns.

Training effectiveness analysis is commonly used to determine whether training programs achieve their intended objectives. Metrics such as completion rate, attendance rate, assessment scores, employee feedback, and post-training performance can be used to evaluate training outcomes. These metrics provide organizations with measurable information about learning activities.

Employee performance and training participation can be influenced by several factors. Previous performance, job role, experience, department, training duration, and attendance may have relationships with training outcomes. Statistical analysis can help identify patterns and relationships within historical employee training datasets.

Machine Learning techniques have increasingly been used in Human Resource Analytics for prediction and classification tasks. Classification algorithms can predict outcomes such as course completion or performance categories, while regression algorithms can estimate numerical assessment scores or performance measures. The quality of predictions depends on data quality, feature selection, and appropriate model evaluation.

Skill-gap analysis is another important area related to employee development. Organizations can compare existing employee skills with required job skills to identify areas where additional training may be useful. Combining skill-gap information with training history can support more targeted learning programs.

Modern employee training platforms increasingly combine learning management, analytics, visualization, and predictive techniques. Interactive dashboards can provide department-wise training performance, course participation, assessment trends, and completion rates. The proposed system follows this approach by integrating employee training analytics with predictive capabilities.

III. SYSTEM ANALYSIS

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The Employee Training Prediction and Analytics system is designed to analyze employee training information and generate meaningful insights. The system accepts information such as employee details, department, job role, training program, training duration, attendance, assessment scores, previous performance, and completion status. The input data is validated to ensure that required fields and supported formats are available. A preprocessing module handles missing values, duplicate records, invalid entries, and inconsistent formats. Numerical and categorical attributes are transformed into suitable formats for analysis. Feature engineering creates useful attributes such as attendance percentage, training duration category, performance category, and completion indicators. The analytics engine calculates training participation, completion rate, average assessment score, attendance rate, and training performance. Department-wise and role-wise analysis identifies differences in employee training outcomes. Training-program analysis compares the effectiveness and participation of different courses. Time-based analysis identifies changes in training participation and performance over different periods. A Machine Learning module can be trained using historical training data to predict selected employee training outcomes. The prediction module can estimate completion probability, assessment performance, or training outcome categories. Visualization components display results through charts, graphs, tables, and KPI cards. The final dashboard provides an integrated view of employee training performance and prediction results.

Existing System

Existing employee training systems primarily focus on course management, employee registration, attendance tracking, training schedules, and completion records. Organizations can use these systems to assign courses and maintain employee learning records. However, many traditional systems provide limited advanced analytics for evaluating training effectiveness. HR teams may need spreadsheets or separate reporting tools to compare training performance across departments and programs. Identifying relationships between attendance, previous performance, and training outcomes may require manual analysis. Historical training trends may not be presented through interactive dashboards. Predictive analysis of employee training outcomes is also limited in basic systems. Training feedback may be stored without detailed analytical processing. Skill gaps may not be automatically identified from available training and performance data. Managers may therefore have difficulty identifying employees who need additional learning support. Data from different training programs may also be stored in separate formats. These limitations create a need for an integrated system that combines training tracking, analytics, visualization, and prediction.

Disadvantages of Existing System

- Limited advanced training analytics.
- Manual comparison of training results.
- Limited prediction of training outcomes.
- Difficult to identify employees requiring additional support.
- Limited department-wise performance analysis.
- Limited training-program comparison.
- Historical trends may be difficult to analyze.
- Training feedback may remain unstructured.
- Limited skill-gap identification.
- Separate tools may be required for advanced analytics.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

- Limited interactive dashboards.
- Manual reporting can consume significant time.

Proposed System

The proposed Employee Training Prediction and Analytics system provides a centralized platform for analyzing employee learning and development data. Users can upload historical employee training datasets or connect the system to an authorized data source. The validation module checks required fields, data types, formats, and basic data quality. The preprocessing module removes duplicate records and handles missing, invalid, and inconsistent information. Feature engineering creates attributes such as attendance percentage, training duration category, previous performance category, and completion indicators. The analytics module calculates training completion rate, attendance rate, average assessment score, employee participation, and training performance. Department-wise analysis compares training outcomes across different organizational units. Job-role analysis identifies differences in learning performance among employee roles. Training-program analysis compares participation, completion, and assessment results across courses. Time-based analysis identifies changes in training activity and outcomes over different periods. A Machine Learning module can be trained using historical employee data to predict selected training outcomes. The prediction module can classify employees according to expected training performance or estimate completion probability where appropriate data is available. Interactive dashboards display KPIs, charts, trends, tables, and prediction results. The proposed system therefore provides a unified solution for employee training monitoring, analytics, and prediction.

Advantages of Proposed System

- Provides centralized employee training analytics.
- Automates training KPI calculations.
- Tracks training participation and completion.
- Calculates attendance and assessment metrics.
- Supports department-wise analysis.
- Supports job-role-wise performance analysis.
- Enables training-program comparison.
- Identifies historical training trends.
- Supports predictive training analysis.
- Helps identify potential learning-support requirements.
- Provides interactive dashboards.
- Reduces manual reporting effort.

IV. METHODOLOGY

The methodology begins with collecting historical employee training and performance data from authorized organizational datasets. The collected data is validated to check its structure, required attributes, data types, and completeness. Data preprocessing is performed to remove duplicate records and handle missing, invalid, and inconsistent values. Numerical attributes such as assessment scores, attendance, training duration, and performance ratings are standardized when required. Categorical attributes such as department, job role, training type, and completion status are transformed into suitable formats. Feature engineering creates attributes such as attendance percentage, training duration category, performance category, and completion indicators. Exploratory Data Analysis is performed to understand employee participation and training-performance patterns. Training completion,

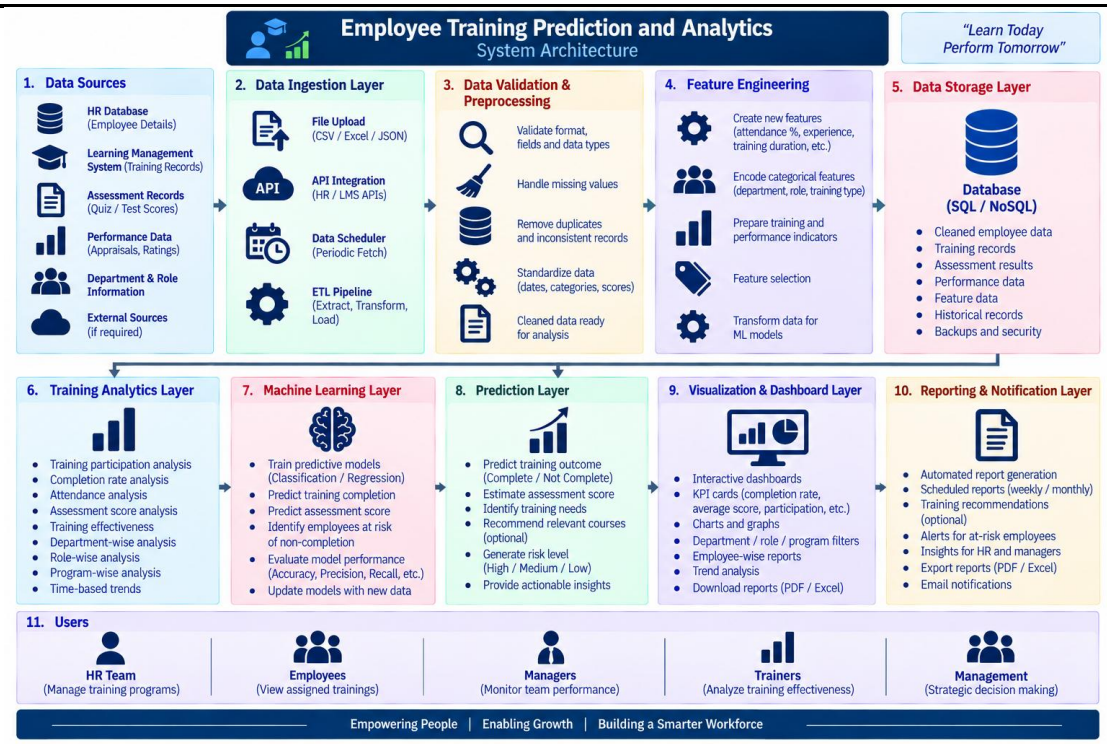
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

attendance, assessment, department, and course-level analyses are performed to identify important trends. Statistical analysis is used to compare training outcomes across different employee groups and programs. Machine Learning models can be trained using suitable historical features to predict training completion, assessment performance, or selected training outcomes. Appropriate evaluation metrics are used to assess the predictive model. The prediction results are combined with analytical insights and displayed through an interactive dashboard.

SYSTEM ARCHITECTURE

The system architecture consists of several interconnected layers that process employee training data and generate analytical and predictive insights. The Data Source Layer provides employee, training, attendance, assessment, performance, department, and job-role information from authorized sources. The Data Input Layer accepts information through CSV files, Excel files, databases, or APIs. The Data Validation Layer checks file formats, required fields, data types, and basic data quality. The Data Preprocessing Layer handles missing values, duplicate records, invalid entries, and inconsistent formats. The Feature Engineering Layer creates attributes such as attendance percentage, training duration category, performance category, and completion indicators. The Data Storage Layer securely stores cleaned employee and training records. The Training Analytics Layer calculates participation, completion rate, attendance rate, assessment scores, and training performance metrics. The Department and Role Analysis Layer compares training outcomes across departments and employee roles. The Training Program Analysis Layer evaluates participation, completion, and performance across different training courses. The Time-Series Analysis Layer identifies changes in training activities and outcomes over different periods. The Machine Learning Layer trains and evaluates models for selected training-related prediction tasks. The Prediction Layer generates estimated training outcomes or performance categories for suitable input data. The Visualization Layer presents results through interactive dashboards, KPI cards, charts, graphs, and tables.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



V. RESULT AND OUTPUT

VI. CONCLUSION

The Employee Training Prediction and Analytics system provides a data-driven platform for monitoring, analyzing, and predicting employee training outcomes. The system converts employee training records into meaningful insights through data preprocessing, statistical analysis, visualization, and Machine Learning techniques.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The system analyzes important factors such as training participation, attendance, assessment scores, training duration, department, job role, previous performance, and completion status. These factors help organizations understand employee learning patterns and evaluate the performance of different training programs.

The interactive dashboard presents information through KPI cards, training completion trends, department-wise analysis, training-status distribution, assessment-score charts, employee performance tables, and training-program analysis. These visualizations make large training datasets easier to understand and support effective monitoring.

The prediction module can estimate an employee's potential training outcome using historical training information. It can identify different levels of completion probability and provide useful insights for identifying employees who may require additional learning support.

In the future, the system can be enhanced with personalized training recommendations, skill-gap analysis, real-time Learning Management System integration, employee feedback sentiment analysis, adaptive learning paths, automated notifications, and advanced predictive models. These enhancements can make the platform more intelligent and useful for improving employee development and organizational training effectiveness.

REFERENCES

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.