International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 10 – Issue 04 || April 2025 || PP. 05-09

Applying Machine Learning to Analyze the Effectiveness of Corporate Training Programs

Murali Krishna Raju Penmetsa
Principal Learning Technology Administrator, CharlesRiver Laboratories
Charlotte, NC, USA

Abstract: The article examines the application of machine learning methods for evaluating the effectiveness of training programs in the corporate sector, driven by the fact that in a rapidly changing market, organizations face the need for objective assessments of their educational processes, requiring appropriate approaches. The purpose of the article is to review existing machine learning models that can identify relationships between employee training and performance metrics. To achieve this objective, the author analyzed classification methods, regression models, and neural networks to process data on training outcomes, employee performance, and engagement levels.

The methodological foundation includes a comparative review of publicly available publications and practical examples from certain companies, which provide a comprehensive exploration of the topic. The findings presented in the article will be useful for training specialists, HR managers, and leaders interested in optimizing their training programs. The conclusions confirm the necessity of applying machine learning for evaluating and subsequently improving educational processes within organizations.

Keywords: machine learning, corporate training, effectiveness, data analysis, forecasting, employee training, HR analytics.

Introduction

In the context of rapid technological advancement and globalization of the economy, employee training and development have become essential components of organizational management. Corporate programs aim to enhance employee qualifications, improve productivity, and adapt to changes in market conditions. To ensure the efficient allocation of resources for such initiatives, companies face the challenge of assessing their impact on organizational outcomes. Traditional methods, such as surveys or self-assessments, often fail to accurately capture the changes resulting from training programs.

The development of big data technologies enables precise and comprehensive evaluation of the effectiveness of corporate training programs. The application of these technologies in the analysis of corporate learning is driven by the need to create an objective basis for managerial decision-making, which supports the optimal allocation of resources. Algorithms reveal patterns that are not always discernible through traditional analysis, providing new opportunities for improving training processes.

The purpose of this article is to examine existing machine learning models that can identify relationships between employee training and their performance metrics.

Materials and Methods

Sharma V. and Jain S. [1] employed support vector machines and decision trees to analyze the success of management training programs. Beinicke A. and Kyndt E. [4] proposed approaches aimed at enhancing the effectiveness of both online courses and traditional training sessions. Studies by Turgeon S. and Lanovaz M. J. [8], as well as Chaturvedi J., Rai M., and Chaturvedi R. [10], demonstrate how machine learning (ML) identifies factors determining training success, including participants' behavioral aspects.

Xiao W. and Hu J. [3] presented a framework for interpreting ML models in online education, which can be adapted for corporate training. Nakhipova V., Suleymenova L., and Adylbekova E. [5] explored methods for predicting academic performance, which are applicable to analyzing corporate training programs.

De Souza B. V. et al. [7] focused on predicting training effectiveness and providing methods for evaluating the outcomes of educational programs. Rafikova K. [9] applied machine learning to analyze social changes, offering insights into training process dynamics within corporate environments. Ahuja R. et al. [11] examined how machine learning supports the evaluation of teamwork effectiveness, which is relevant for corporate collaboration development programs.

Tamang M. D. et al. [6] noted that ML applications help understand the impact of training programs on corporate outcomes. Malchenko P. et al. [2] proposed methods for assessing companies' growth prospects using ML, which is crucial for evaluating the business impact of training programs. Zhang W. [12] introduced a

machine learning model for predicting financial risks, aiding in the economic feasibility analysis of training programs.

Additionally, platforms such as edstellar and educate-me [13, 14] provided data on changes in corporate training from 2019 to 2024. Sources such as allsee.team [15] and www.unite.ai [16] offered examples of companies like AT&T and Gradescope leveraging machine learning to analyze corporate training program effectiveness.

Thus, studies on the application of machine learning to analyze corporate training effectiveness offer diverse approaches. However, unresolved issues remain, such as a lack of studies comparing results from different methods using identical datasets and insufficient attention to long-term training program effects and the influence of corporate culture on outcomes.

The methodological basis of this research included a comparative analysis of publicly available publications and practical examples from companies, enabling a comprehensive examination of the chosen topic.

Results and Discussion

Statistical data indicate that the use of machine learning to evaluate corporate training programs has become a prominent development trend between 2019 and 2024. The data presented in Figure 1 confirm its growing role [13, 14].

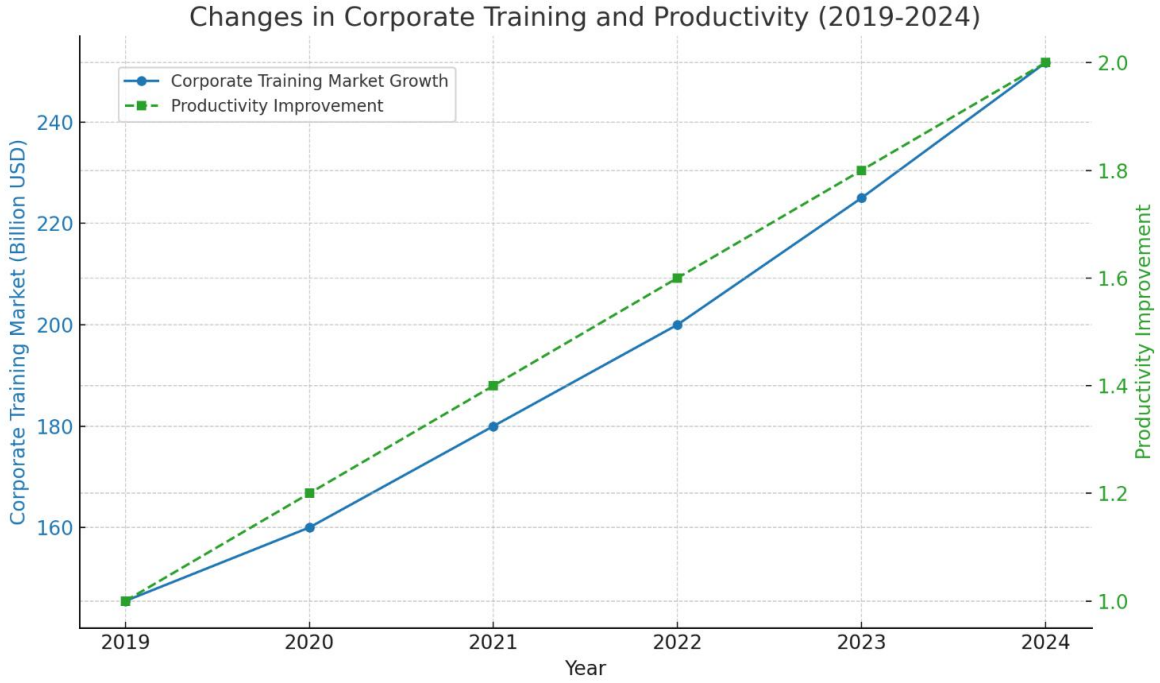

Fig.1. Changes in corporate training for 2019–2024 [13–14].

Various data sources are utilized to analyze the effectiveness of training programs, including quantitative metrics, textual feedback, video recordings of training sessions, and sensor data. The choice of model depends on the specific task. For predicting changes in employee performance based on historical data, ensemble methods such as gradient boosting and random forests are applied. For clustering employees to identify groups with shared characteristics, density-based algorithms like DBSCAN (Density-Based Spatial Clustering of Applications with Noise) are used. These algorithms identify cohorts that traditional methods may overlook. For analyzing textual feedback, including uncovering emotional and semantic patterns, natural language processing methods such as transformers, including BERT (Bidirectional Encoder Representations from Transformers), are employed. These models reveal hidden connections within textual data. In some cases, a combination of different models is applied to account for multiple factors, thereby minimizing the risk of underestimating relationships between features [1, 2, 6, 12]. The evaluation of algorithm performance is based on several criteria, illustrated below in Figure 2.

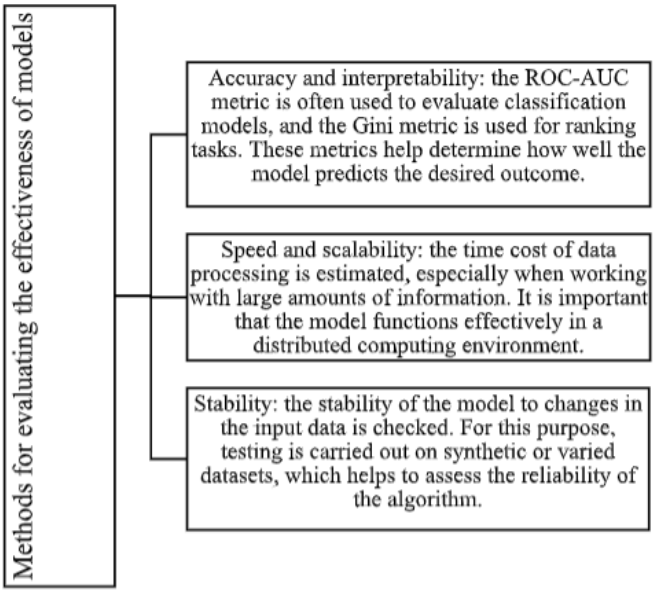

Fig.2. Methods for evaluating the effectiveness of models

The following section examines the specifics of applying machine learning to analyze the effectiveness of corporate training. The implementation process begins with collecting data from various systems, such as LMS (Learning Management System), HR (Human Resources) systems, and employee surveys. These systems provide metrics on engagement, test results, and performance indicators before and after training. At this stage, data cleaning is conducted, which includes removing duplicates and standardizing formats.

Subsequently, data preparation involves creating new features, such as calculating intervals between training modules or extracting keywords from textual feedback using natural language processing methods. These additional features enhance the accuracy of models. Tools such as Python with libraries like pandas, NumPy, Scikit-learn, TensorFlow, and specialized visualization platforms such as Power BI and Tableau are utilized for data processing [3, 4, 8].

Modeling is performed using various algorithms. Classification and regression methods, such as logistic regression, decision trees, Random Forest, and Gradient Boosting, are applied to predict employee success or course completion likelihood. Clustering methods, such as K-Means or DBSCAN, are used to segment employees for personalized recommendations. Text analysis employs natural language processing techniques like TF-IDF, Word2Vec, or transformers such as BERT.

For large datasets, distributed computing platforms like Apache Spark or Hadoop are utilized. Data storage relies on systems such as Amazon Redshift, Snowflake, or Google BigQuery. Data processing is handled with ETL (Extract, Transform, Load) tools, including Apache NiFi or Airflow.

Time-series data, such as changes in employee performance post-training, are analyzed using models like ARIMA (AutoRegressive Integrated Moving Average), LSTM (Long Short-Term Memory Networks), or Prophet. These models help track performance changes over time. Causal analysis methods, such as A/B testing or causal graphs, are used to identify causal relationships between training and performance improvements [5, 7, 10].

After deploying models, their performance must be monitored and updated as needed. MLOps tools like MLflow, Kubeflow, and TFX are used to automate deployment and monitoring processes. These tools enable models to be retrained with new data. Additionally, dashboards are created to visualize key metrics, such as ROI (Return on Investment) from training programs or the percentage of employees completing courses [2, 9, 11].

The following section provides examples of machine learning applications used to analyze the effectiveness of corporate training programs.

AT&T employs artificial intelligence and machine learning technologies to enhance user experience and improve employee efficiency. Specifically, ML is utilized to evaluate and plan employee work, contributing to effective customer service and increased productivity. According to reports, AT&T's productivity increased by 5% [15].

Gradescope, a platform combining machine learning technologies, simplifies the grading process. This approach saves instructors time and ensures objective knowledge assessment, which is beneficial for corporate training programs [16].

Advancements in ML methods have also opened new possibilities for data analysis. For example, graph models can be used to study employee network connections, while generative neural networks can create synthetic data for hypothesis testing. Table 1 below outlines promising directions for applying machine learning methods to analyze the effectiveness of corporate training programs.

Table 1. Promising areas of application of machine learning methods in analyzing the effectiveness of corporate training programs (compiled by the author).

Area	Description	Application Examples	Implementation Features	Role in Work Processes
Learning Outcome Prediction	Use of algorithms to analyze employees' performance based on their characteristics and training.	Study of productivity improvement post-training, assessment of successful project involvement probability.	Requires data on employees' previous experience to build predictions.	Aids in deciding the feasibility of training programs for specific employees and optimizing training plans.
Training Sensitivity Analysis	Analysis of the relationship between various training factors and outcomes to identify effective elements.	Use of machine learning to determine factors impacting task performance.	Requires substantial data and subsequent model adjustment.	Helps enhance existing training programs and adapt them to employee and company needs.
Employee Classification by Effectiveness	Grouping employees based on their post-training performance.	Categorizing employees into successful, average, and those needing additional training.	Requires accurate data labeling and determination.	Enhances the personalization of training programs and optimizes resource allocation.
Personalized Learning Paths	Development of tailored training programs using algorithms that consider individual characteristics.	Creation of courses based on recommendations from recommender systems.	Requires integration with learning platforms and employee data analysis.	Facilitates effective training, improves motivation, and increases engagement.
Training Quality Assessment via Text Analysis	Use of natural language processing to analyze feedback and evaluations of training courses.	Automatic analysis to identify hidden issues in training materials.	Requires data annotation for training models and challenges in result interpretation.	Enhances material comprehension.
Engagement Assessment through Gamification	Use of machine learning to evaluate employee engagement in gamified training processes.	Analysis of game metrics to measure engagement in learning activities.	Requires adaptation of gamification systems to employee needs and LMS integration.	Increases employee engagement, making training processes engaging and personalized.
Training Time Optimization	Application of machine learning to identify optimal timing and format for employee training.	Optimization of session duration and frequency to enhance material retention.	Complexity in evaluating the impact of time and format, requiring precise algorithm tuning.	Improves training efficiency, minimizing time and costs.

Thus, the application of machine learning in evaluating the effectiveness of corporate training programs facilitates a shift from intuitive approaches to analytically grounded solutions. Identifying hidden patterns, optimizing training modules, and forecasting long-term outcomes represent only a fraction of the potential unlocked by ML.

Conclusion

In conclusion, the application of machine learning methods for analyzing the effectiveness of corporate training programs has been examined. The findings demonstrate that the use of classification and regression algorithms improves the accuracy of evaluating the impact of training initiatives on metrics such as employee productivity, engagement, and satisfaction levels.

One of the key conclusions is the ability of these methods to identify hidden relationships between training parameters and performance outcomes—relationships that traditional tools are unable to detect. Machine learning models have proven effective in predicting the outcomes of training programs, providing organizations with tools for data-driven decision-making in training management.

The practical significance of the obtained data lies in the ability to accurately plan and adapt training programs based on objective indicators. The use of such data facilitates the optimization of training expenses and the improvement of performance metrics. Implementing these technologies enhances the transparency of evaluation processes.

References

[1]. Sharma V., Jain S. Managers Training Programs Effectiveness Evaluation by using different Machine Learning Approaches //2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). – IEEE, 2020. – pp. 1453-1457.
[2]. Malchenko P. et al. Application of machine learning algorithms in determining the value perspectives of corporations //E3S Web of Conferences. – EDP Sciences, 2024. – Vol. 508. – pp. 03006.
[3]. Xiao W., Hu J. Analyzing Effective Factors of Online Learning Performance by Interpreting Machine Learning Models //IEEE Access. – 2023. – Vol. 11. – pp. 132435-132447.
[4]. Beinicke A., Kyndt E. Evidence-based actions for maximising training effectiveness in corporate E-learning and classroom training //Studies in Continuing Education. – 2020. – Vol. 42 (2). – pp. 256-276.
[5]. Nakhipova V., Suleymenova L., Adylbekova E. Determining student performance using machine learning methods //ILIM. – 2024. – Vol. 41 (3). – pp. 5-20.
[6]. Tamang M. D. et al. Improving business intelligence through machine learning algorithms //2021 2nd International Conference on Intelligent Engineering and Management (ICIEM). – IEEE, 2021. – pp. 63-68.
[7]. De Souza B. V. et al. Analyzing and predicting Overall Equipment Effectiveness in manufacturing industries using machine learning //2022 IEEE International Systems Conference (SysCon). – IEEE, 2022. – pp. 1-8.
[8]. Turgeon S., Lanovaz M. J. Tutorial: Applying machine learning in behavioral research //Perspectives on Behavior Science. – 2020. – Vol. 43 (4). – pp. 697-723.
[9]. Rafikova, K. (2024). Predicting Social Change Using Machine Learning Algorithms // Теория и практика общественного развития. – 2024. – Vol. 4. - pp. 90–96.
[10]. Chaturvedi J., Rai M., Chaturvedi R. Effectiveness of training program on “design & analysis of machines using MATLAB & motionview” //IETE Journal of Research. – 2021. – Vol. 67 (5). – pp. 620-633.
[11]. Ahuja R. et al. Machine learning and student performance in teams //Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21. – Springer International Publishing, 2020. – pp. 301-305.
[12]. Zhang W. Research on Corporate Financial Risk Prediction Model Based on Machine Learning //2024 IEEE 2nd International Conference on Control, Electronics and Computer Technology (ICCECT). – IEEE, 2024. – pp. 535-540.
[13]. Corporate Training Market Insights for 2025. [Electronic resource] Access mode: https://www.edstellar.com/blog/corporate-training-market-analysis (accessed: 01.01.2025).
[14]. 10 Main Corporate Training Trends + Stats [2024]. [Electronic resource] Access mode: https://www.educate-me.co/blog/corporate-training-trends (accessed: 01.01.2025).
[15]. 10 лучших примеров использования AI и ML. [Electronic resource] Access mode: https://allsee.team/ai-and-ml-use-cases?utm_source (accessed: 01.01.2025).
[16]. 10 лучших инструментов ИИ для образования. [Electronic resource] Access mode: https://www.unite.ai/ru/10-лучших-инструментов-искусственного-интеллекта-для-образования/?utm_source (accessed: 01.01.2025).