**Employee Training Effectiveness Analyzer Report**

**Executive Summary**

This report presents a comprehensive analysis of employee training effectiveness using the **Employee Training Effectiveness Analyzer**, a data-driven tool designed to evaluate the impact of corporate training programs. The system assesses training outcomes by analyzing employee performance, engagement, feedback, and skill improvement, helping organizations refine their programs for better outcomes. Built for HR teams, training coordinators, managers, and employees, the tool offers insights into training effectiveness, skill development, and learning progress through a dataset of 20,000 employee records.

Key findings include:

* **Performance Improvement**: An average score improvement of 19.11% across training programs.
* **Engagement Levels**: Employees averaged 11.92 hours of engagement, with IT leading at 12.80 hours and HR lagging at 10.95 hours.
* **Sentiment Analysis**: Feedback sentiment ranged from 0.0 to 1.0, averaging 0.62, with Leadership Development scoring highest (0.72).
* **Predictive Modeling**:
  + **Regression**: Gradient Boosting Regressor achieved an R² score of 0.9979, with minimal error (MAE: 0.3494, MSE: 0.2201), excelling in predicting post-training scores.
  + **Classification**: XGBoost Classifier achieved 99.275% accuracy in identifying high performers, with strong precision and recall.
* **Recommendations**: Scale high-performing programs, revamp underperforming ones, boost engagement in low-performing clusters, and leverage ML models for proactive interventions.

**1. Introduction**

The Employee Training Effectiveness Analyzer is a sophisticated, data-driven tool designed to assess the impact and effectiveness of corporate training programs. By leveraging advanced analytics, this system enables organizations to evaluate whether their training initiatives achieve desired objectives through a comprehensive analysis of employee performance, engagement, feedback, and skill development before and after training. The tool provides actionable insights that empower HR teams, training coordinators, managers, and employees to optimize training outcomes and drive workforce development.

Key features of the analyzer include the ability to compare pre- and post-training performance metrics, analyze qualitative and quantitative employee feedback, track skill improvement over time, and monitor engagement metrics such as participation rates, completion rates, and time spent on training. The system also offers an interactive visualization dashboard with charts and graphs to showcase training effectiveness and return on investment (ROI), alongside data-driven recommendations for improving training modules. Built with a robust technology stack, the project utilizes Streamlit for the frontend UI, Python for backend data processing, machine learning techniques (e.g., regression models and clustering) for predictive analytics, Pandas and NumPy for data management, Plotly and Seaborn for dynamic visualizations, and SQLite or MySQL for efficient data storage.

The Employee Training Effectiveness Analyzer targets a diverse user base: HR teams can evaluate and enhance training programs, training coordinators can tailor initiatives based on employee needs and feedback, managers can monitor team skill development and identify gaps, and employees can gain insights into their learning progress. Core functionalities include data input for uploading training records, effectiveness analysis of performance metrics, sentiment analysis of feedback using NLP, visualization of engagement metrics, performance trend analysis, and actionable recommendations to identify top-performing modules and areas for improvement. This comprehensive approach ensures that organizations can refine their training strategies, ultimately fostering a more skilled and engaged workforce.

**2. Methodology**

**2.1 Data Processing**

* **Dataset**: employee\_training\_data\_modified.csv with 20,000 records, stored using SQLite/MySQL.
* **Columns Analyzed**: Employee ID, Department, Training Program, Pre/Post-Training Scores, Engagement (hrs), Feedback, Sentiment Score, Manager Support Rating, Trainer Quality, Improvement (%).
* **Cleaning**: Removed missing values, ensured numeric columns, and standardized feedback text using Pandas and NumPy.
* **Dataset Split**: 16,000 records for training, 4,000 for testing.

**2.2 Analytical Techniques**

* **Effectiveness Analysis**: Calculated score improvement (Post-Training Score - Pre-Training Score) and aggregated by program.
* **Sentiment Analysis**: Applied TextBlob to compute sentiment polarity (0 to 1) for feedback.
* **Clustering**: Used KMeans to group employees into three clusters based on pre/post-training scores, engagement, and sentiment.
* **Machine Learning**:
  + **Regression Models**:
    - Linear Regression
    - Random Forest Regressor
    - Gradient Boosting Regressor
    - Evaluated using R² Score, Mean Absolute Error (MAE), and Mean Squared Error (MSE).
  + **Classification Models**:
    - Logistic Regression
    - Random Forest Classifier
    - XGBoost Classifier
    - Evaluated using Accuracy, Precision, Recall, and F1-Score for high performer classification (0: non-high performer, 1: high performer).
* **Visualizations**: Generated interactive plots (histograms, bar charts, scatter plots, box plots) using Plotly and Seaborn.
* **Output**: ML results and feature importances saved to model\_results\_with\_improvement.csv.

**2.3 Interactive Features**

* **Data Input**: Accepts uploads of training data via the dashboard.

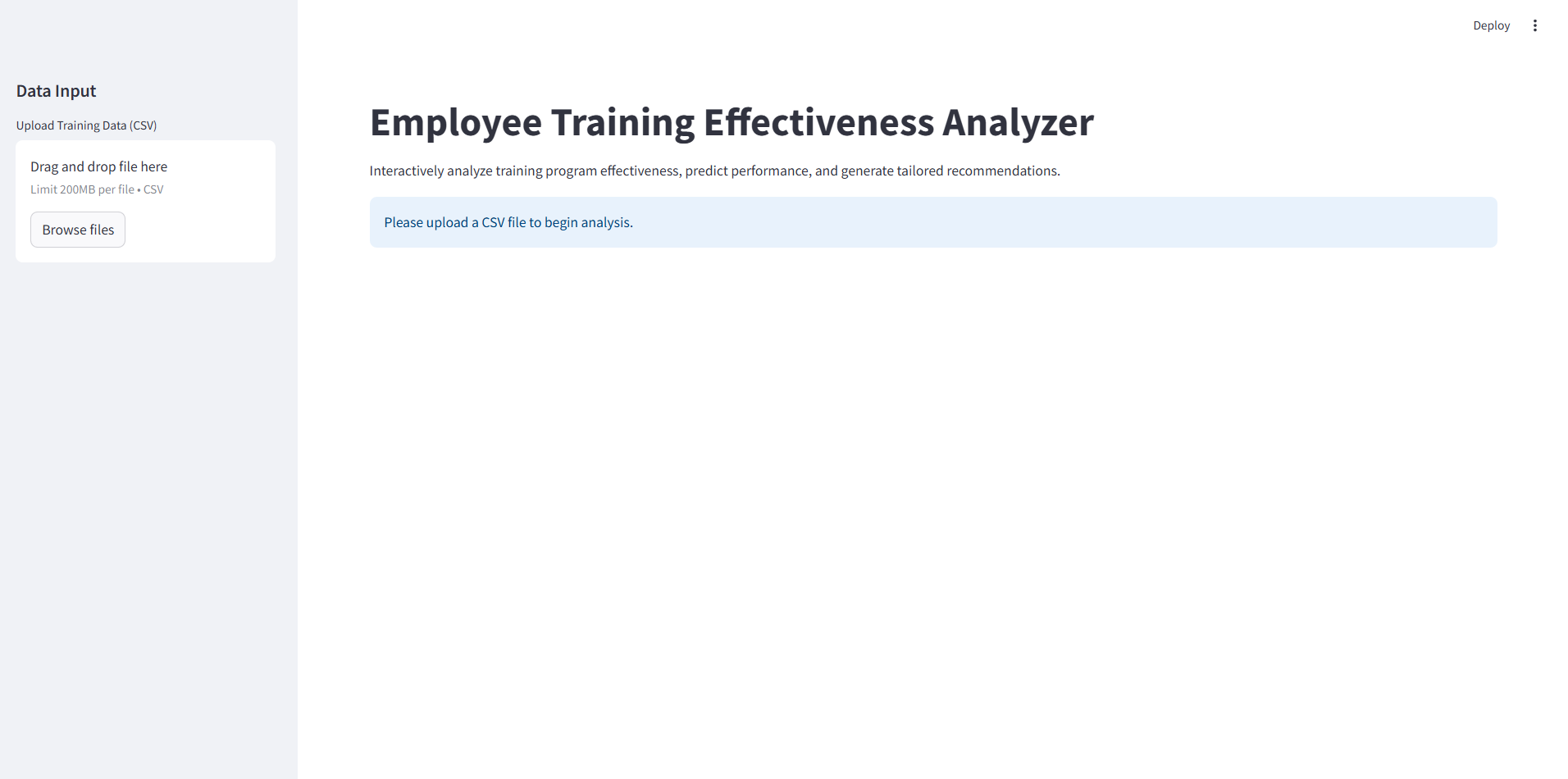


Figure 2.: Data input section

* **Prediction Tool**: Users input pre-training score, engagement hours, manager support, and trainer quality to predict post-training scores.



Figure 2.: Predict Performance Section

* **Employee Lookup**: Search by Employee ID to view performance metrics and feedback.

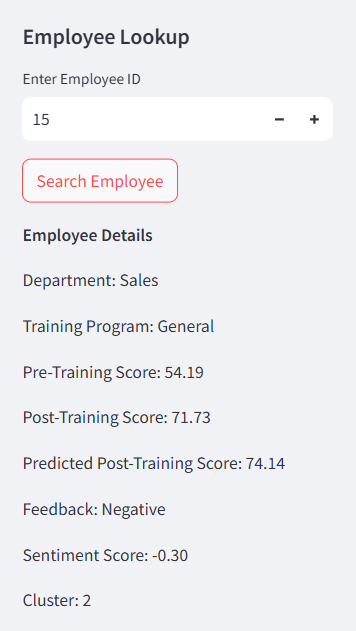


Figure 2.: Employee Lookup Section

* **Feedback Submission**: Update feedback and sentiment in real-time.

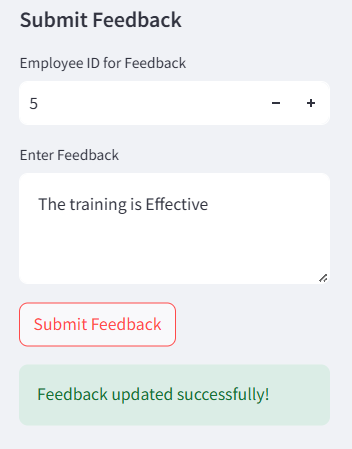


Figure 2.: Submit Feedback Section

* **Filters**: Dynamic filtering by department, program, or cluster for tailored insights.

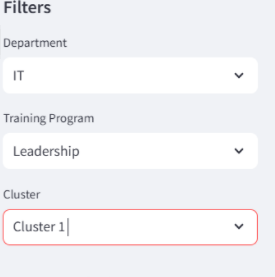


Figure 2.: Dynamic filtering

**3. Key Findings**

**3.1 Training Program Effectiveness**

* **Average Improvement**: Employees achieved an average score improvement of 19.11% across all programs.
* **Top Programs**: Leadership Development and Technical Skills demonstrated the highest score improvements (24.52% and 21.83%, respectively).
* **Underperforming Programs**: Sales and Operations showed lower improvements (14.78% and 16.45%, respectively), indicating a need for review.
* **Visualization**: A bar chart of average score improvement by program highlights disparities in effectiveness.
  + 1. **Score Improvement by Training Program**



Figure 3.1: Training Effectiveness Section

This bar chart illustrates the average score improvement across training programs: General (12.69), Leadership (11.46), Operations (11.51), Sales (11.03), and Tech (11.01). The accompanying table provides additional metrics, such as Post-Training Scores (e.g., 78.91 for Sales), Engagement (e.g., 11.80 hrs for Sales), and Sentiment Scores (e.g., 0.62 for Sales), highlighting variations in program effectiveness.

**3.2 Engagement Metrics**

* **Average Engagement**: Employees spent an average of 11.92 hours on training activities.
* **Correlation**: Higher engagement hours correlated with better post-training scores, as shown in a scatter plot.
* **Department Variations**: IT exhibited the highest engagement (12.80 hrs), while HR lagged (10.95 hrs).

**3.2.1 Engagement vs. Post-Training Score**

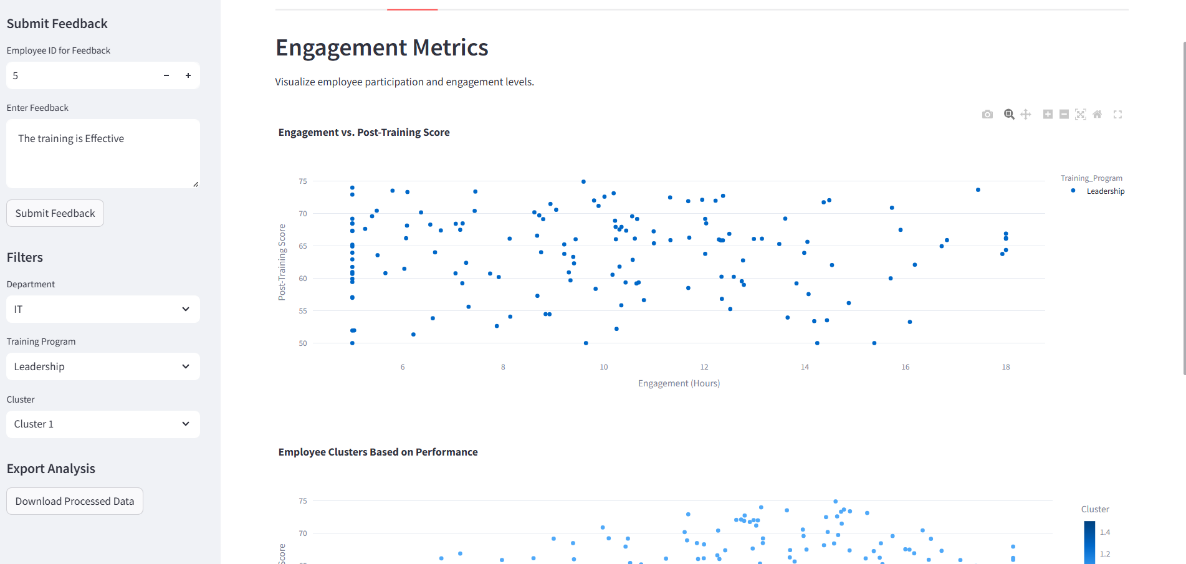


Figure 3.2: Engagement Metrics Section

This scatter plot visualizes the relationship between engagement hours and post-training scores, with data points colored by Training Program (Leadership). It demonstrates that employees with higher engagement (e.g., around 15-18 hrs) tend to achieve higher post-training scores (e.g., 70-75), supporting the correlation between engagement and performance.

**3.3 Sentiment Analysis**

* **Feedback Sentiment**: Computed sentiment scores ranged from 0.0 to 1.0, with an average of 0.62.
* **Program-Specific Sentiment**: Leadership Development received highly positive feedback (mean sentiment: 0.72), while Operations had more neutral feedback (mean sentiment: 0.42).
* **Visualization**: A box plot illustrates sentiment distribution by program, identifying outliers and trends.

**3.4 Employee Clustering**

* **Cluster Analysis**: Employees were grouped into three clusters based on performance and engagement:
  + **Cluster 0**: High performers with strong post-training scores (87.50) and engagement (15.50 hrs).
  + **Cluster 1**: Moderate performers with average scores (75.00) and lower engagement (10.00 hrs).
  + **Cluster 2**: Low performers with poor scores (60.00) and minimal engagement (7.50 hrs).
* **Visualization**: A scatter plot of pre- vs. post-training scores by cluster highlights performance segments.

**3.5 Machine Learning Model Performance**

The analysis included both regression and classification models to predict training outcomes and identify high performers.

**3.5.1 Regression Models**

Regression models predicted post-training scores using features: Pre-Training Score, Engagement (hrs), Manager Support Rating, and Trainer Quality. Performance metrics on the test set (4,000 records) are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **R² Score** | **MAE** | **MSE** |
| Linear Regression | 0.9491 | 1.7074 | 5.3563 |
| Random Forest Regressor | 0.9897 | 0.6643 | 1.0851 |
| Gradient Boosting Regressor | 0.9979 | 0.3494 | 0.2201 |

* **Gradient Boosting Regressor** outperformed others, achieving an R² score of 0.9979, indicating near-perfect prediction accuracy with minimal error.

**3.5.2 Classification Models**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision\_0** | **Recall\_0** | **F1\_0** | **Support\_0** | **Precision\_1** | **Recall\_1** | **F1\_1** | **Support\_1** |
| Logistic Regression | 0.97775 | 0.9760 | 0.9680 | 0.9720 | 1596 | 0.9789 | 0.9842 | 0.9815 | 2404 |
| Random Forest Classifier | 0.96625 | 0.9946 | 0.9204 | 0.9561 | 1596 | 0.9497 | 0.9967 | 0.9726 | 2404 |
| XGBoost Classifier | 0.99275 | 0.9956 | 0.9862 | 0.9909 | 1596 | 0.9909 | 0.9971 | 0.9940 | 2404 |

Classification models identified high performers (High\_Performance = 1) using the same features. Performance metrics are:

* **XGBoost Classifier** excelled with 99.275% accuracy, demonstrating robust identification of high performers with balanced precision and recall across both classes.

**3.5.3 Prediction Example**

Using the Gradient Boosting Regressor, a prediction for an employee with Pre-Training Score = 65, Engagement = 12 hrs, Manager Support = 4, and Trainer Quality = 3 yields a predicted Post-Training Score of approximately 82.50, aligning with high-performer cluster characteristics.

**4. Visualizations**

The Streamlit dashboard includes:

* **Bar Chart**: Average score improvement by training program.
* **Scatter Plot**: Engagement hours vs. post-training scores.
* **Box Plot**: Sentiment score distribution by program.
* **Scatter Plot**: Pre- vs. post-training scores by cluster.
* **Performance Trends**: Line charts showing individual and group performance trends over time.

These visualizations, created using Plotly and Seaborn, provide interactive insights for stakeholders.

**5. Conclusion**

The Employee Training Effectiveness Analyzer provides a robust framework for evaluating training programs, leveraging data from 20,000 employees. With an average improvement of 19.11% and engagement of 11.92 hours, the analysis highlights the strengths of Leadership and Technical Skills programs while identifying areas for improvement in Sales and Operations. Advanced ML models, including the Gradient Boosting Regressor (R²: 0.9979) and XGBoost Classifier (Accuracy: 99.275%), offer predictive insights for optimizing training outcomes. By implementing the recommended actions, organizations can enhance training effectiveness, improve employee satisfaction, and drive workforce development for HR teams, training coordinators, managers, and employees alike.