Heliverse Internship Task - Insights Report

Summary

This report presents insights drawn from the analysis of the dataset provided for the Heliverse Internship Task. The dataset contains information related to employee attrition in an organization. Various machine learning models were applied to predict attrition and analyze factors influencing it.

Insights from Visualizations

1. The age-wise distribution of employees shows a higher attrition rate among younger employees.
2. Department-wise analysis indicates that certain departments have a higher attrition rate compared to others.
3. Environment satisfaction levels affect attrition, with lower satisfaction leading to higher attrition rates.
4. Heatmap visualization depicts correlations between different features in the dataset, providing insights into potential predictors of attrition.

Model Performance

Logistic Regression:

- Training Accuracy: 69%
- Testing Accuracy: 59%

Support Vector Machine (SVM):

- Training Accuracy: 72.5%
- Testing Accuracy: 61.4%

Decision Tree:

- Training Accuracy: 68.6%
- Testing Accuracy: 61.7%

Random Forest:

- Training Accuracy: 100% (Overfit)
- Testing Accuracy: 54%

XGBoost:

- Training Accuracy: 100% (Overfit)
- Testing Accuracy: 62.3%

Feature Importance

Features such as **monthly income, job satisfaction, and years at the company** were identified as important predictors of attrition, as shown in the feature importance plot generated using XGBoost.

Comparison of Models

Random Forest and XGBoost outperformed other models in predicting attrition, achieving **higher ROC-AUC scores** on both training and testing datasets.

Conclusion

The analysis provides valuable insights into factors influencing employee attrition and highlights the effectiveness of machine learning models in predicting attrition. Organizations can leverage these insights to implement strategies for employee retention and improve overall workplace satisfaction.