Capstone Project Proposal: Predicting Employee Attrition in Large Organizations

1. Introduction

Employee turnover in large organizations leads to increased recruitment costs, reduced morale, and loss of institutional knowledge. By leveraging machine learning techniques, HR departments can proactively identify employees at risk of leaving and implement targeted retention strategies.

2. Problem Statement

High employee attrition negatively impacts productivity and organizational stability. Predicting which employees are likely to leave will allow HR teams to take preventive action, thereby improving employee retention and reducing costs.

3. Dataset

The **IBM HR Analytics Employee Attrition and Performance Dataset** (available on Kaggle) contains features such as:

- **Demographics:** Age, gender, education, marital status
- **Work-related factors:** Job satisfaction, work-life balance, department, performance rating
- **Compensation & benefits:** Salary, promotion history, years at company

4. Methodology

To predict employee attrition, the following machine learning techniques will be explored:

- **Data Preprocessing:** Handling missing values, normalizing features, addressing class imbalance
- **Exploratory Data Analysis (EDA):** Identifying trends and correlations in employee retention
- **Model Selection:** Evaluating logistic regression, decision trees, random forests, and neural networks
- **Hyperparameter Tuning:** Optimizing model performance

5. Evaluation Metrics

The effectiveness of the model will be assessed using:

- **Accuracy:** Overall correctness of predictions
- **Precision & Recall:** Measuring false positives and false negatives

- **F1 Score:** Balancing precision and recall
- **AUC-ROC Curve:** Evaluating classification performance

6. Expected Challenges & Mitigation Strategies

- **Data Imbalance:** If attrition cases are disproportionately low, resampling techniques will be applied.
- **Bias in Predictions:** Careful feature selection and bias mitigation strategies will ensure fairness.
- **Interpretability:** Using SHAP values to understand model decision-making and improve trustworthiness.

7. Conclusion

This project aims to provide a practical, data-driven solution to the challenge of employee attrition. By predicting which employees are at risk of leaving and understanding why, HR teams can take proactive steps to improve retention. The final model will not only support decision-making but also offer valuable insights that can enhance employee satisfaction and organizational stability.