**Artificial Intelligence and**

**Machine Learning**

Project Report Semester-IV (Batch-2022)

Title of the Project:

# Prediction of Employee Retention



**Supervised By: Submitted By:**

Mr. Rajeev Bhardwaj Suyash Kr. Sharma (2210991000)

Shubham Goyal (2210990845)

Sujal Valecha (2210990868)

**Department of Computer Science and Engineering**

**Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab**

**PROBLEM STATEMENT**

Build a Model to Predict Employee Retention

Employee retention is a significant concern for organizations across industries due to its direct impact on productivity, morale, and operational costs. High turnover rates can result in the loss of valuable talent, disrupt workflow continuity, and necessitate expensive recruitment and training processes. To address this challenge, there is a need to develop an effective predictive model using artificial intelligence and machine learning techniques to forecast employee retention probabilities accurately.

The problem at hand involves building a predictive model that can analyse various factors contributing to employee turnover and provide insights into which employees are at a higher risk of leaving the organization. This model aims to utilize historical data on employee demographics, performance metrics, job satisfaction levels, tenure, and other relevant attributes to predict the likelihood of an employee resigning from their position.

**SCOPE OF PROJECT**

The scope of the project encompasses several key aspects related to building a predictive model for employee retention using AI/ML techniques. The following outlines the specific areas within the scope of this endeavour:

1. Data Collection: Gather relevant data pertaining to employee demographics, performance metrics, job satisfaction scores, tenure, and other pertinent factors influencing retention. Ensure data integrity, accuracy, and compliance with privacy regulations during the collection process.
2. Data Preprocessing: Cleanse and preprocess the collected data to handle missing values, outliers, and inconsistencies. Perform data transformation, normalization, and encoding as necessary to prepare the dataset for model training.
3. Feature Selection and Engineering: Identify and select informative features that have a significant impact on employee retention. Conduct feature engineering to create new variables or derive meaningful insights from existing ones to enhance the predictive power of the model.
4. Model Development: Implement and train machine learning algorithms such as RandomForestClassifier and XGBClassifier on the prepared dataset. Explore various model architectures, hyperparameters, and optimization techniques to develop accurate and robust predictive models.
5. Model Evaluation: Assess the performance of the trained models using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Conduct cross-validation and validation on a separate test dataset to evaluate model generalization and reliability.
6. Interpretability and Insights: Analyse the trained models to gain insights into the factors contributing to employee retention. Interpret model predictions and identify key drivers influencing employee turnover within the organization.

the project aims to contribute to the overall success and sustainability of organizations by optimizing workforce management practices, fostering employee loyalty, and creating a conducive environment for growth and innovation.

**DATA OVERVIEW**

Datatype and Source

<https://www.kaggle.com/code/farzadnekouei/employee-turnover-prediction-xgboost-vs-catboost/comments>

**DATA PREPROCESSING STEPS**

* Data Cleaning: Handle missing values (imputation or removal).

Address inconsistencies and errors (format corrections, outlier detection).

* Feature Engineering: Create informative features from existing data (e.g., time-based features, ratios, aggregations).

Explore additional features from user behavior or network data

* Data Splitting: Divide data into training, validation, and testing sets for model building and evaluation.

**Algorithm Selection:**

1. Random Forest Classifier:
   * Random Forest is a versatile ensemble learning algorithm that combines multiple decision trees to make predictions. It is robust to overfitting, handles both numerical and categorical data well, and provides feature importance scores, which can be insightful for understanding the factors

influencing employee retention.

1. XGBoost Classifier:
   * XGBoost (Extreme Gradient Boosting) is a powerful gradient boosting algorithm known for its efficiency and performance. It sequentially builds a series of decision trees, each correcting the errors of the previous one. XGBoost often yields highly accurate predictions and is suitable for large datasets

**conclusion**

In conclusion, the project to predict employee retention using machine learning techniques offers valuable insights and solutions for organizations striving to mitigate the challenges associated with high turnover rates. By leveraging algorithms such as Random Forest Classifier and XGBoost Classifier, we have developed predictive models capable of forecasting employee retention probabilities accurately.

**THANK YOU**