

Proactive attrition system

- **Project Description:** The core purpose of this project is to address the significant and recurring financial and operational risk posed by employee attrition the persistent and often sudden loss of talent. High turnover drives up direct costs (recruitment and training) while simultaneously causing a critical erosion of institutional expertise and team morale. To counteract this, the primary goal is to shift the company's approach to human capital management from reactive to proactive by developing a Proactive Attrition Risk System. This system will leverage Machine Learning (ML) and predictive analytics to solve the fundamental challenge: the inability of traditional HR methods to quickly and accurately forecast which employees are most likely to leave. By accurately identifying high-risk individuals before their departure, the system enables the timely implementation of targeted retention strategies.
- **Group Members**
 1. Ahmed Fekry Mohamed (**TEAM LEADER**)
 2. Yousif Moaz ELbadry
 3. Mazen Kamal Morsy
 4. Shaimaa Ahmed fouad
 5. Youssef Ahmed fouad
- **Project Objectives**
 1. Discovering causes that can tell us why an may employee leave his job
 2. Forecast the chance for an employee to turnover his job
 3. Proactively deal with the attrition and prevent talent lose

- **Tools & Technologies**
 1. IDE: GOOGLE COLAB, ANACONDA
 2. Programming languages and libraries: Python and some of its advanced libraries like pandas, numpy, matplotlib, seaborn, sklearn, xgboost, scipy, imblearn
 3. ML algorithms: Logistic Regression, Random Forest, XGB, K-NN
- **Milestones**
 1. Milestone 1: Data Collection, Exploration, and Preprocessing (**finished**)
 2. Milestone 2: Advanced Data Analysis and Feature Engineering (**finished**)
 3. Milestone 3: Machine Learning Model Development and Optimization (**under progress**)
 4. Milestone 4: MLOps, Deployment, and Monitoring (**under progress**)
- **KPIs (Key Performance Indicators):** Please specify the key metrics for measuring the Success of your project based on the following aspects.
 1. *Data Quality*
 - Percentage of missing values handled 0%
 - Data accuracy after preprocessing: 99%
 - Dataset diversity (representation of different categories): 32%
 2. *Model Performance*
 - Model accuracy (Accuracy/F1-Score): 96% / 91% (class 1)
 - Model prediction speed (Latency): 1 milliseconds
 - Error rate (False Positive/False Negative Rate): 0.8% / 10%