**Name: Pham Quoc Trung** - 范国忠 (Tee) **- Student ID:** 1137168

**Course Code:** CM763E - Artificial Intelligence **- Instructor:** Qazi Mazhar Ul Haq (卡其)

**Project:** Employee Attrition Prediction Using Survival Analysis & Machine Learning

## 1. Project Summary

This project presents a time-to-event (survival) machine learning pipeline designed to predict employee attrition using IBM HR Analytics data. By formulating it as a survival analysis problem, we implemented and evaluated interpretable models, including Cox Proportional Hazards and Random Survival Forest, alongside an XGBoost classifier. SHAP explainability identified key attrition drivers such as overtime, stock options, and job tenure. The final output provides HR with a scalable tool for generating individualized risk scores and survival probabilities, facilitating proactive retention strategies and data-driven decision-making.

***Github Repository Link:*** https://url-shortener.me/664I

## 2. Execution instructions

Step 1: Clone the repository and ensure Python 3.11+ is installed.

Step 2: Create a virtual environment

```
python -m venv venv
source venv/bin/activate  # On Windows: venv\Scripts\activate
```

Step 3: Install dependencies using pip and the provided requirements.txt file:

```
pip install -r requirements.txt
```

Step 4: Run the main analysis pipeline:

```
python sourcecode/hr_survival.py
```

Step 5: Check outputs

```
jupyter notebook notebooks/analysis.ipynb
```

## 3. Special Notes

- The IBM HR dataset used is synthetic; model performance may vary with real-world organizational data.
- Key libraries include lifelines, scikit-survival, xgboost, and shap. Ensure they are correctly installed.
- All outputs, including risk tables, SHAP summary plots, and performance metrics, are saved to the outputs/ directory.
- If you encounter version conflicts, please refer to the specific library versions listed in requirements.txt.