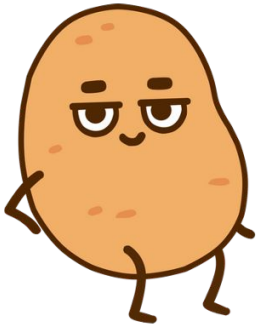
THE UNIVERSITY OF MELBOURNE

Master of Data Science Industry Project Meeting

Predictive Analytics for Personnel Separation

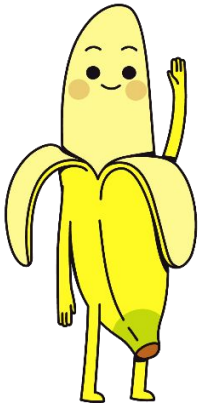
MAST90106 & MAST90107 Group 20
17/10/2024

Team Members



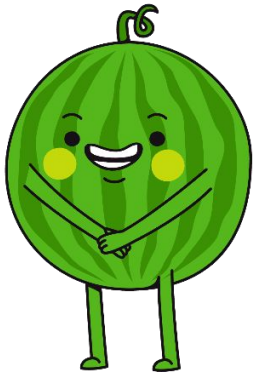
Xuan Ji

- Recording and summarizing meeting minutes
- Create the Analytical Dashboard



Junheng Yu

- Team leader
- Development the system



Pengjiabei Tang

- Communications and emails within the team
- Models implement
- Result dashboard

Yufeng Liu

- Initial data cleaning and visualization
- Create the Analytical Dashboard

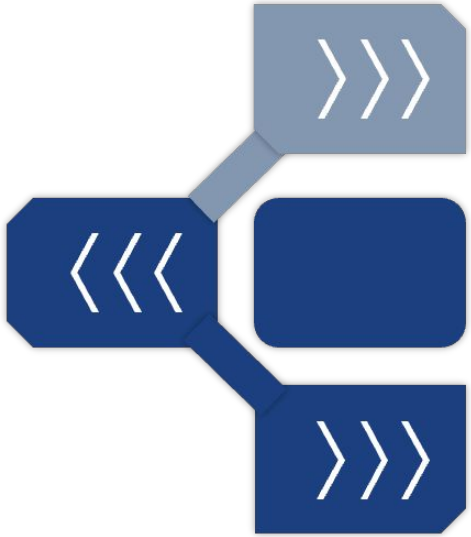
Cheng Qian

- Preparing slide and agenda for meeting
- Create the Analytical Dashboard

Project Overview

Background

- **Employee turnover** is costly, affecting productivity and increasing recruitment costs.
- **Traditional methods** often react too late and miss complex factors like job satisfaction and compensation.
- **Data-driven solutions** can help predict risks and enable proactive retention strategies.



Host

C4 Engineering Pty Ltd

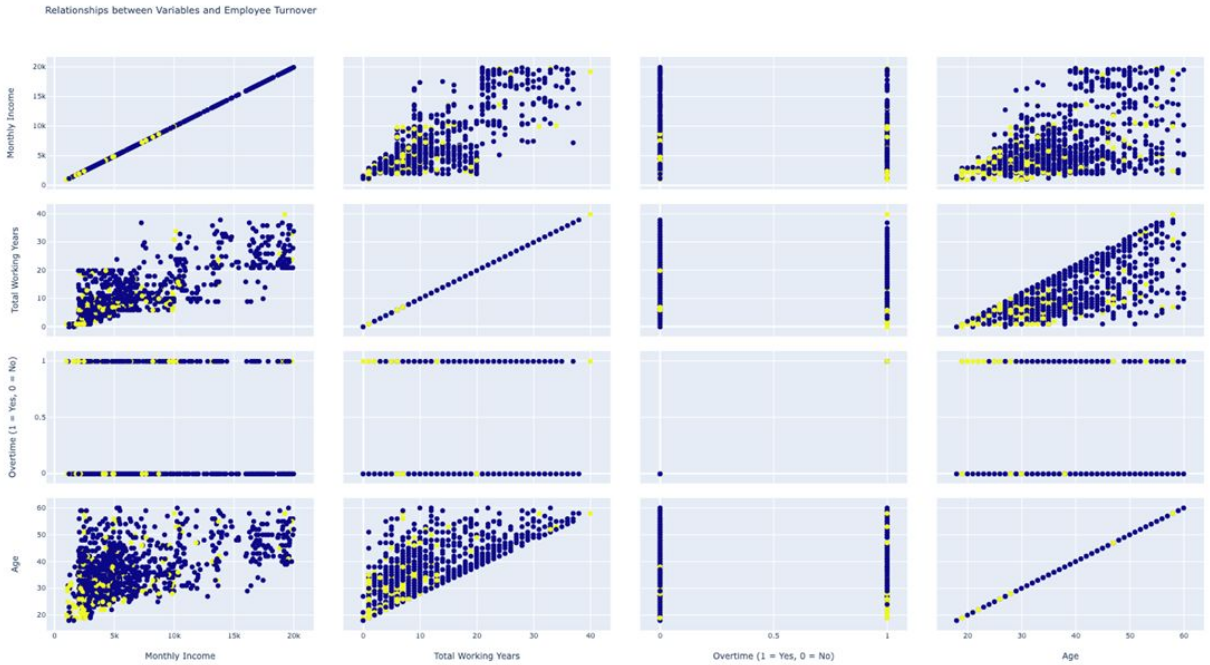
Goal

- Identify key factors contributing to employee turnover.
- Develop predictive models to identify high-risk employees.
- Build a system for real-time visualization of personnel data.

01

Data

First View of the Data— Kaggle Data



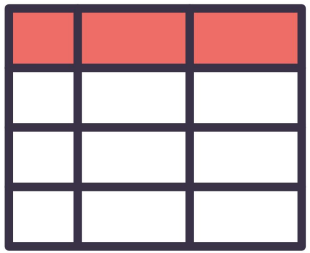
- Monthly Income
- Overtime
- Total Working Years
- Age

Data Generation

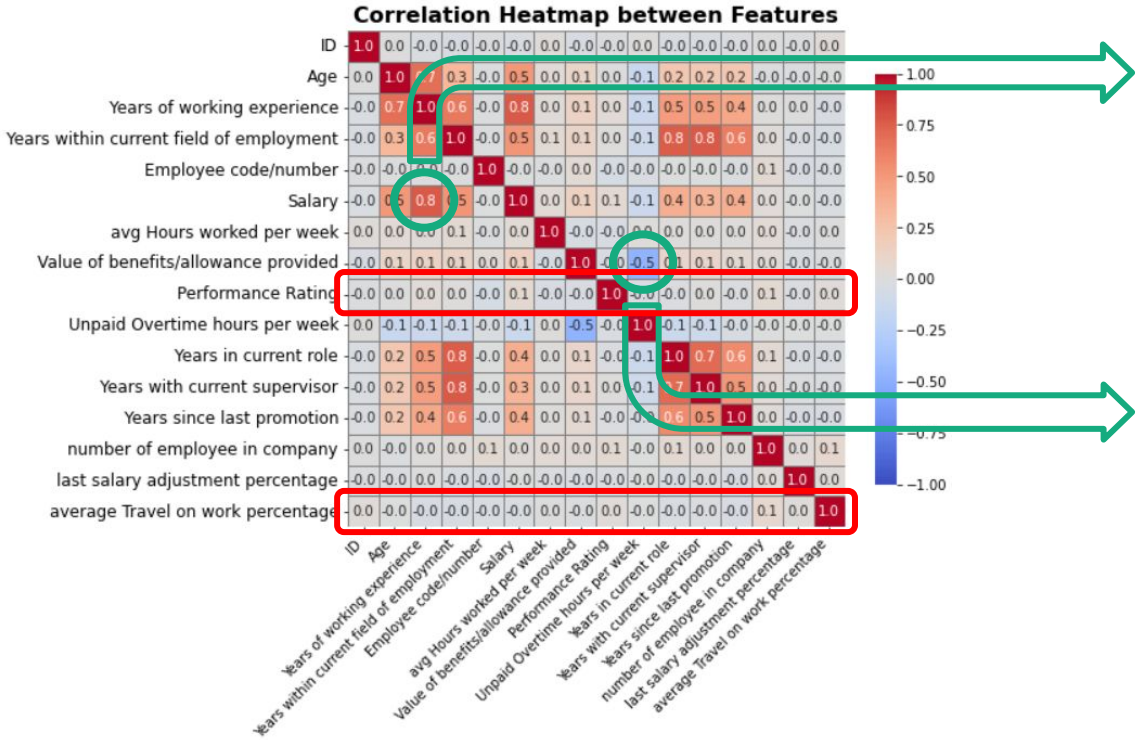
- Dataset lacked key features for business needs.
- Generated new features based on data relationships and assumptions.
- Used AI to create features
- Built a more complete dataset for model training.

Kaggle Dataset

Synthetic Features



Relationship Between Features



High positive correlation between Salary and Years within current field of employment

High negative correlation between Value of benefits/allowance and Unpaid Overtime hours

02

Traditional ML Models

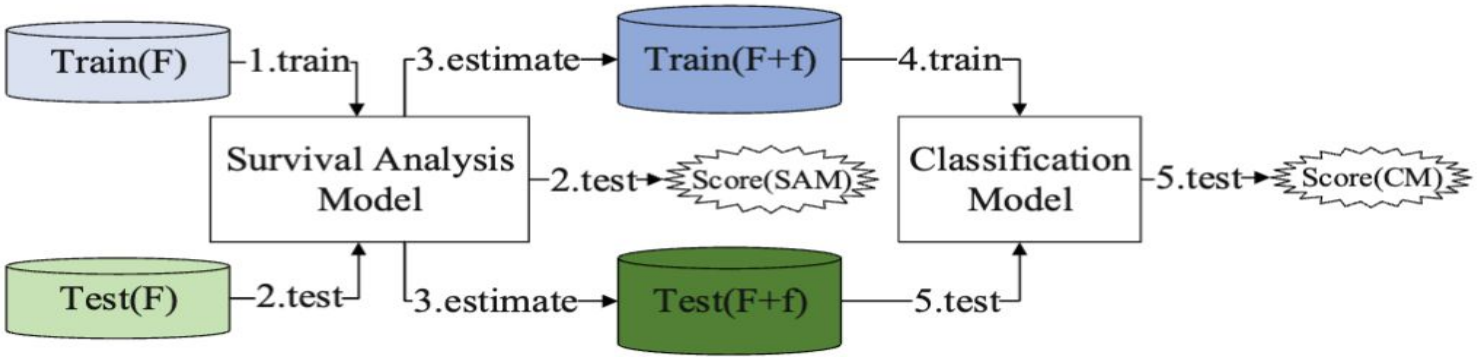
Traditional ML Models

Models	Accuracy on Kaggle Data	Accuracy on Synthetic Data	Recall on Label 1 (Synthetic Data)
Logistic Regression	0.87	0.97	0.89
Naive Bayes	0.77	0.95	0.84
Support Vector Machines	0.85	0.95	0.80
XGBoost	0.82	**0.96**	0.87
Random Forest	0.90	**0.96**	0.84
RFRSF Hybrid Model (New)	0.92	**0.97**	**0.93**

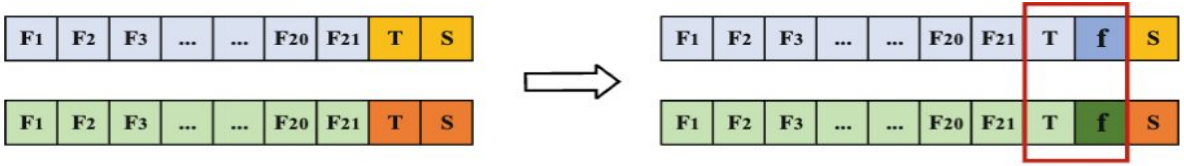
03

New Hybrid Model

Introduce New Hybrid Model – Framework



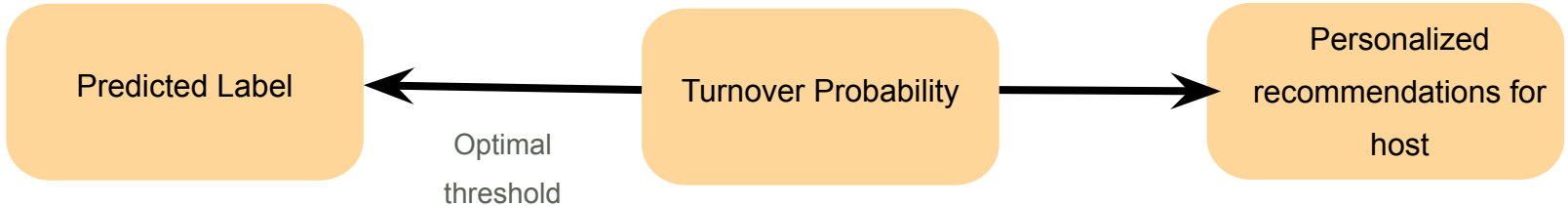
in Data table



- F1-F21...: Original features;
- T: survival time (Years in the company);
- f: Survival rate;
- S: label (0 for stay, 1 for leave);

Ref: Jin, Ziwei, et al. "RFRSF: Employee turnover prediction based on random forests and survival analysis."

Introduce New Hybrid Model – Results



0: p<=0.288, Predicted to stay in the company

1: p>0.288, Predicted to leave the company

0-1 possibility of leaving in 1 Year

High risk: p>0.432
"60% probability of leaving within 1 years. Immediate retention measures recommended."

Medium risk: 0.288<=p<=0.432
"30% probability of leaving within 1 years. Close monitoring and preventive measures advised."

Low risk: p<0.288
"15% probability of leaving within 1 years. Regular monitoring necessary."

Advantages of New Hybrid Model

THE UNIVERSITY OF MELBOURNE

Enhancing Prediction

Combines Random Forest with survival analysis, enhancing prediction.

Time-related Feature

Uses 'years at company' to understand tenure impact on turnover.

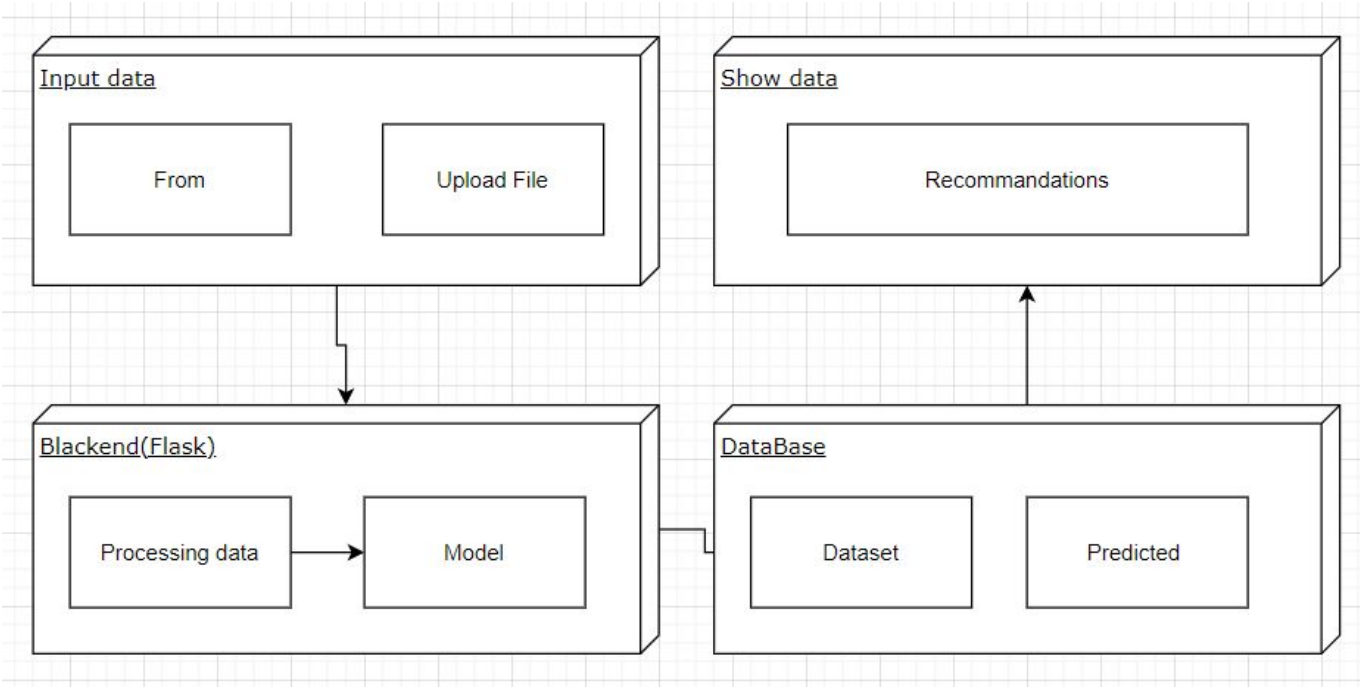
Personal recommendation

Provides leaving probability for next 1 year, enabling targeted retention strategies.

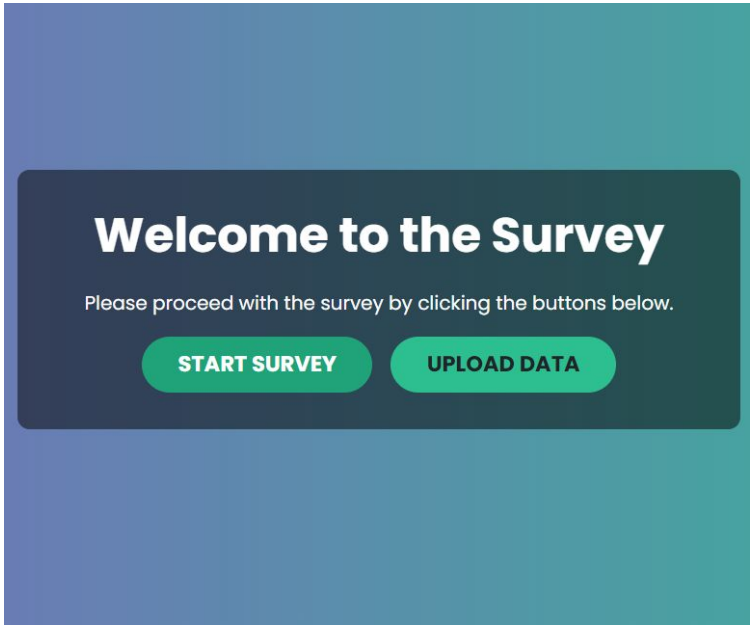
04

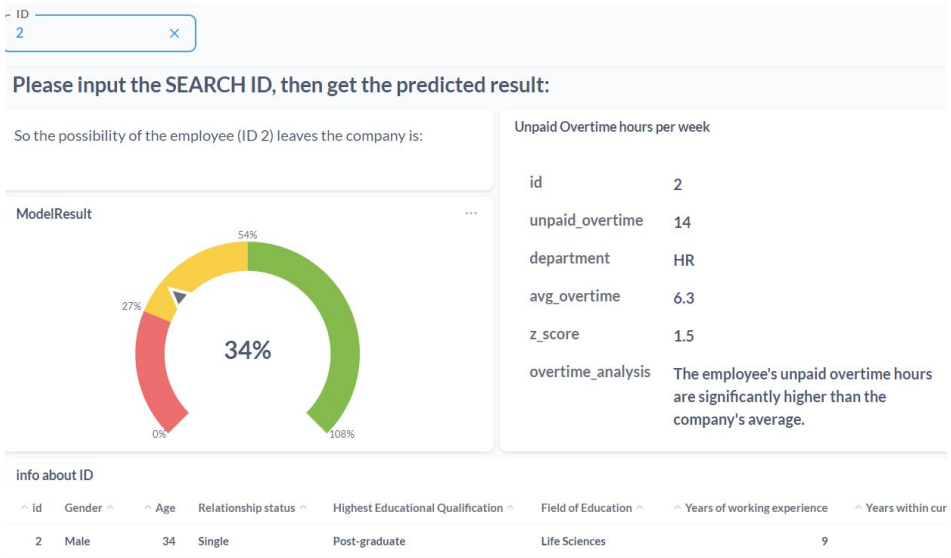
System Introduction

Introduction



Demonstration





05

Conclusion

Data and Final Model Limitation

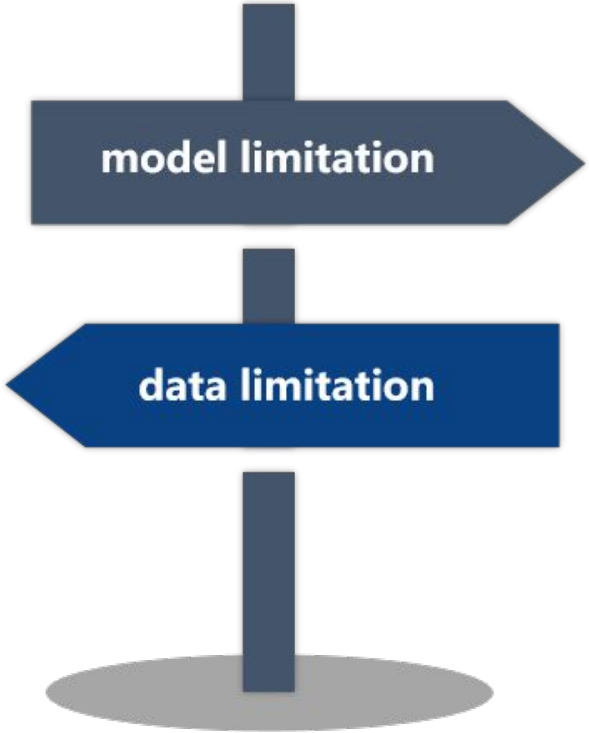
Lack of Real-World Complexity:

1. Generated data lacks real-world complexity.

2. May lead to poor performance on actual data.

Overlooking Key Edge Cases:

1. May overlook important edge cases.
2. Reduce the model's generalization ability.



Overfitting Risk:

1. Very high accuracy scores
2. May not generalize well to new data

Time Limitation:

1.Based on static features
2.May miss dynamic changes in employee/company situations

Conclusion

- **Overall**
  - Kaggle datasets
  - Initial models
  - Generate data
  - Final model
  - Dashboards
- **Future work**
  - use the result to see how can we decrease the turnover rate and what the company to do to make the employee to stay longer.
  - Anticipate personnel movement to implement measures

THE UNIVERSITY OF MELBOURNE

Thank you