Retirement prediction using machine learning

Nishi, Yoshihiro

AGENDA

1. RETIREMENT ISSUES
2. HR DATA VISUALIZATION
3. ANALYSIS OF HR DATA USING MACHINE LEARNING

1

RETIREMENT ISSUES

The impact of an increase in sudden leavers

An increase in leaving workers has a big negative impact on companies.

- Cost increase for recruiting employees
  - Interview cost
  - Job ad cost
- Increase in employee education expenses
  - Cost of new employee education
- It becomes difficult to form a business strategy systematically

The impact of an increase in sudden leavers

- Cost increase for recruiting employees
  - Interview cost
  - Job ad cost
- Increase in employee education expenses
  - Cost of new employee education
- It becomes difficult to form a business strategy systematically

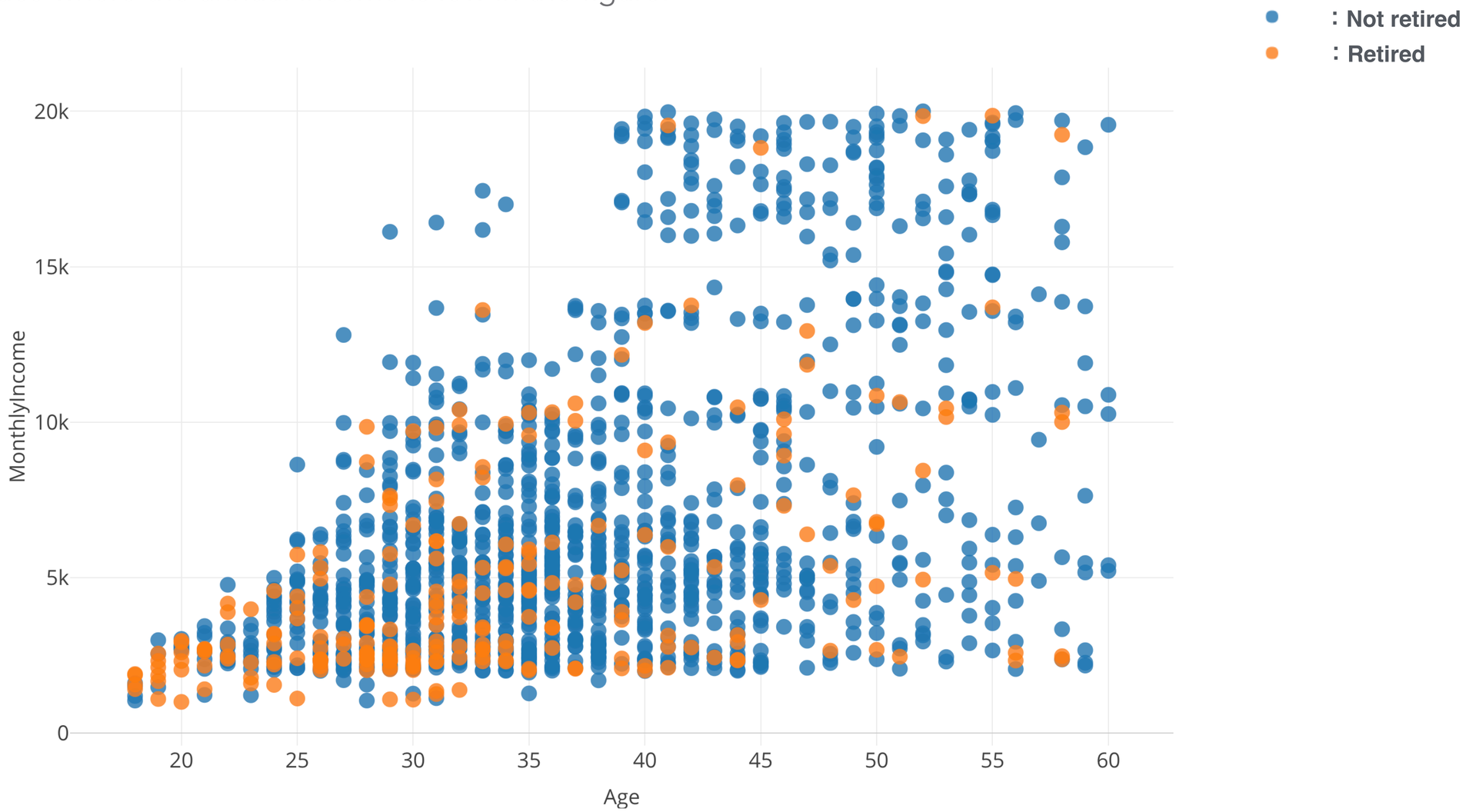
Risks can be avoided in advance by using machine learning to predict retirement and taking steps to prevent retirement

2

HR DATA VISUALIZATION

Scatter plot of annual income and age, color-coded according to retirement status

Younger generations have lower annual income, and it can be seen that they tend to increase as their age increases.
Although the volume zone is near the lower left of the scatter plot, it can be seen that a large number of leaving workers are gathered at the place where the annual income is low in the figure.



3

ANALYSIS OF HR DATA USING MACHINE LEARNING

Calculation of Importance Using Random Forest

The objective variable is "presence or absence of retirement". It can be said that OverTime (presence or absence of overtime work) and JobRole (business position) help to identify the leaver.

	MeanDecreaseAccuracy
OverTime	**22.20**
JobRole	**11.00**
StockOptionLevel	**8.11**
JobSatisfaction	5.77
MaritalStatus	5.69
TotalWorkingYears	4.95
Age	4.86
MonthlyIncome	4.77
JobInvolvement	4.74

	MeanDecreaseGini
JobRole	**17.43**
MonthlyIncome	**16.87**
OverTime	**16.35**
Age	13.18
DailyRate	12.13
EmployeeNumber	11.12
YearsAtCompany	10.29
MonthlyRate	10.19
TotalWorkingYears	10.10

＊上位10変数のみを抜粋

Details of analysis of turnover rate using decision tree

The prediction accuracy of the decision tree analysis is 88.53%, which is a high value. Also, in decision tree analysis, it can be said that MonthlyIncome (monthly income) and OverTime (presence or absence of overtime work) are useful for identification of leaving workers.

Prediction result

Prediction accuracy : Correct rate 88.53%

		Attrition	
		No	Yes
prediction	No	1201	137
	Yes	32	100

Importance of each variable

MonthlyIncome	OverTime
28.26	24.08
TotalWorkingYears	**JobRole**
18.58	16.89
StockOptionLevel	**DailyRate**
11.29	10.59
Age	**MaritalStatus**
9.46	9.10
HourlyRate	**YearsInCurrentRole**
8.06	7.95

Retirement prevention process using HR data

By collecting and analyzing HR data such as employee information and attendance data, and implementing measures to prevent turnover based on the analysis results, it is possible to raise the accuracy of the measures and prevent retirement more effectively.