### **Attrition Analysis**

A Case Study of IBM's Attrition Prediction

Xuan Wang Jingr

Yi Nian Jingrou Wei



### Introduction

IBM HR Analytics Employee Attrition & Performance

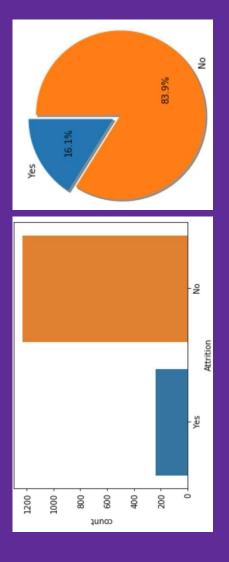
Predict attrition of your valuable employees

#### **Motivation**

Construct correlations between factors
Analyze attrition distribution of features
Implement and evaluate models to predict the attrition rate
Find best plan to minimize attrition rate

### Data Visualization

;	:	;	i	į	:
EmployeeNumber	<b>*</b>	2	4	2065	2068
EmployeeCount	~		_	1	-
EducationField	Life Sciences	Life Sciences	Other	Medical	Medical
Education	2	· <del>-</del>	2	က	8
Department DistanceFromHome Education EducationField EmployeeCount EmployeeNumber	~	ω	2	2	8
Department	Sales	Research & Development	Research & Development	Sales	Research & Development
DailyRate	1102	279	1373	1023	628
Age Attrition BusinessTravel DailyRate	Yes Travel_Rarely	No Travel_Frequently	Travel_Rarely	No Travel_Frequently	Travel_Rarely
Attrition	Yes	o N	Yes	N <sub>o</sub>	N
Age	41	49	37	49	34
	0	~	2	1468	1469

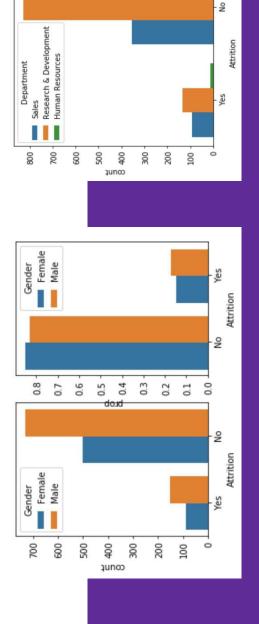


Attrition Total: 237 Attrition Rate: 16.1% (high)

# **Attrition Distribution Analysis**

Comparisons of attrition amount and attrition rate in 'Gender' and 'Department' Research & Development

0.8



prop 0.4

0.2

9.0

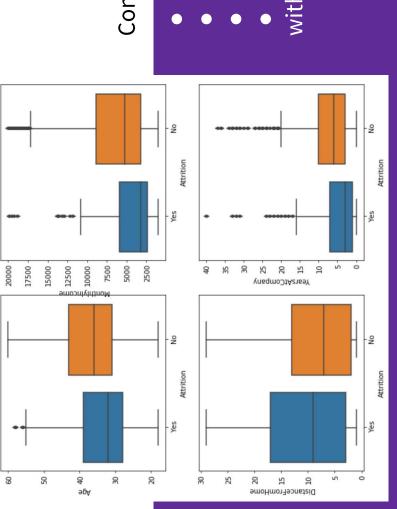


Yes

0.0



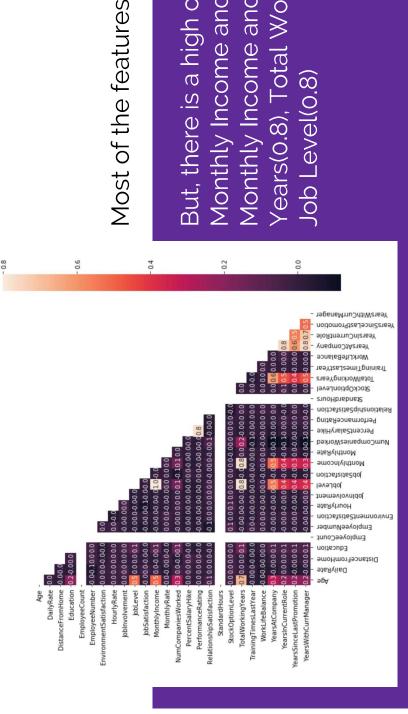
# **Attrition Distribution Analysis**



### Comparisons of attrition in

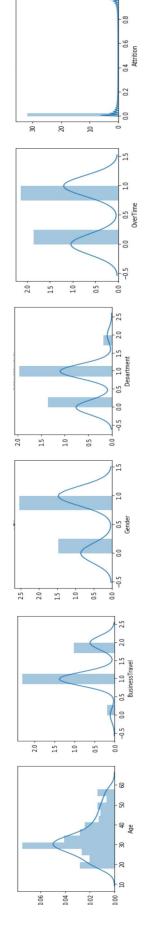
- , 'Age',
- 'Monthly Income',
- 'Distance From Home',
  - 'Years at Company' with Box Plot.

# **Correlations between Features**



Most of the features are uncorrelated

But, there is a high correlation between Years(0.8), Total Working Years and Monthly Income and Job Level(1.0), Monthly Income and Total Working



Add second order or third Try to fit a linear model. order terms to features appeared non-linear.

**EnvironmentSatisfaction** DistanceFromHome JobSatisfaction BusinessTravel JobInvolvement MaritalStatus MonthlyIncome Department Education JobLevel Gender

RelationshipSatisfaction YearsSinceLastPromotion YearsWithCurrManager YearsInCurrentRole PerformanceRating **TotalWorkingYears** StockOptionLevel YearsAtCompany JobLevelsquare Edusquare NumCompaniesWorked

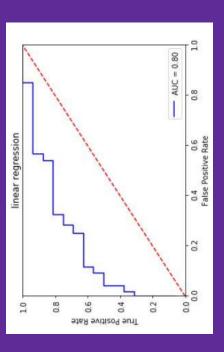
OverTime

OLS Regression Results

Dep. Variable:	Attrition	Attrition R-squared:	0.346	Afi
Model:	ST0	Adj. R-squared:	0.335	
Method:	Least Squares	F-statistic:	31.90	<u>S</u>
Date:	Tue, 27 Nov 2018	Prob (F-statistic):	2.50e-115	) -
Time:	17:09:04	Log-Likelihood:	-7201.7	≷
No. Observations:	1470	AIC:	1.445e+04	(
Df Residuals:	1446	BIC:	1.458e+04	34
Df Model:	24			
Covariance Type:	nonrobust			

After comparing AIC and significance test, a linear model with terms shows on the left with 34.6% R square level.

ROC = 0.80 Looks good but relations between features and attrition may be more complicated...



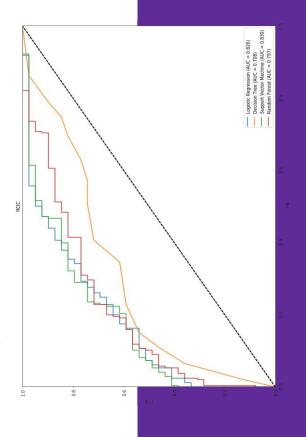
#### Data Cleaning

0	0.446350	3.446350 0.742527	-1.010909	-0.891688	-0.660531	1.383138	0.379672	1.458650	-0.918921
-	1.322365	-1.297775	-0.147150	-1.868426	0.254625	-0.240677	-1.026167	-0.685565	1.088232
2	0.008343	0.008343 1.414363	-0.887515	-0.891688	1.169781	1.284725	-1.026167	1.458650	-0.918921

l _	0	_
1.591746	-0.628241	1.591746
-1.591746	0.628241	-0.918921 -1.591746
-0.918921	1.088232	-0.918921
1.458650	-0.685565	1.458650
	1.458650 -0.918921 -1.591746 1.591746 1	

- Check if there exists missing data
- Make the attrition column numeric, 1 for 'yes' and 0 for 'no'
- Remove variables that are the same all the time and not meaningful
- Assign an unique indicator for each categorical variables in the dataset
  - Make all the features in our dataset have mean 0 and std 1 for future convenience

Logistic Regression vs. Decision Tree vs. Support Vector Machine vs. Random Forest



### **Models Comparison**

Compared by ROC:

Logistic Regression: 0,826 Decision Tree: 0,708

SVM: 0.830

Random Forest: 0.797

## Feature Importance

Mutual Information

Feature Importance from Random Forest

		Monthlyland
thiylncome 0.033729928271344134 0.033729928271344134 0.033052820732690114 0.033052820732690114 0.033052820732690114 0.00240661985597114 0.0024145828896673383 0.0024145828896673383 0.0024145828996673383 0.0024145828996673383 0.0024145828996673383 0.0024145828996673383 0.0024145828996673383 0.0024145828996673383 0.0024145828996673383 0.00244598248924848364649852 0.0024466198528763476 0.0024466198528763476 0.0024466198528763476 0.002446619852 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.002446649853 0.00244640737589650 0.00244698964660 0.0024824839566073 0.00248224956 0.00248122104730996 0.00248122104730976674303 0.00248248188985 0.00248122104730097674303 0.002482429556 0.00248122104730097674303 0.002482429556 0.00248122104730097674303 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002482428566738 0.00248122104730097674818981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481248981221243 0.002481221243 0.002481221243 0.002481221243 0.002481221243 0.002481248981221243 0.002481221243 0.002	0	Age
thlyIncome 0.033729928271344134  Time Yes indicator 0.02140661985559714  SinCurrentRole 0.0214682896673383  Time No indicator 0.02148289667383  Time No indicator 0.0213280417582972  Time No indicator 0.013168485084734677  TalStatus Divorced indicator 0.019168688688498488  IngitimestalstYear 0.01973703576334765  CationField Technical Degree indicator 0.01095329310183954  IngitimestalstYear 0.01095329310183954  Inestatus SingitimestalstYear 0.01095329310183954  Inestatus SingitimestalstYear 0.0009949488395666073  Ole Healthcare Representative indicator 0.009949488395666073  Ole Realthcare Representative indicator 0.009949488395666073  Ole Manufacturing Director indicator 0.009849488395666073  Ole Manufacturing Director indicator 0.00988114718298966  Ole Sales Executive indicator 0.004812210475048395  Indistatus Maricel indicator 0.004812210475048395  Ole Research Director indicator 0.004812210475048395  Indicator 0.004812210475048395  Ole Research Director indicator 0.004812210475048395  Ole Research Director indicator 0.0038041188981221243  Ole Research Director indicator 0.003804188981221243  Ole Research Director indicator 0.0048804417827466  Ole Research Director indicator 0.004880441888818787	0.045058039119793	TotalMorking
Time Yes indicator	0.033729928271344134	lotalWorking
Interverse   0.027406619855597114	0.033052820732690114	DallyRate
0.024145828896673383	0.027406619855597114	Monthlykate
0.02169236715204148	0.024145828896673383	YearsAtComp
0.02132804175829972	0.02169236715204148	HourlyRate
Idicator 0.01916488084734677  O019168485084734675  Degree indicator 0.019467963224052392  Degree indicator 0.010447963224052392  at indicator 0.01099494832966073  Everlopment indicator 0.0099949832966073  St indicator 0.0099949832966073  Development indicator 0.0099949832966073  Indicator 0.00838114274798098  Indicator 0.00838114274798098  Indicator 0.0083814427498098  Indicator 0.00785281777189906  Indicator 0.00785281777189906  Indicator 0.00785281777189906  Indicator 0.00785281777189906  Indicator 0.00785286473758687875  Indicator 0.004880714747342355  Indicator 0.004880718981271243  Indicator 0.0042806470654303  Indicator 0.00325128848185896	0.02132804175829972	DistanceFron
volicator         0.018468798646649853           Degree indicator         0.01797703576344765           Degree indicator         0.011686886688498488           0.010447963224052392           astor         0.010493703576371           actor         0.01093728410945356           Development indicator         0.0099372841094356           Development indicator         0.008581548367678237           rector indicator         0.0083114774798098           pindicator         0.007852873777189906           noice and an activity indicator         0.007852873777189906           noice and activity indicator         0.0048121047356985395           tor         0.0048121047375986759333           tor         0.0048127137863838755           for         0.0048127104760483395           for         0.0048127104766781           on         0.004812713766383975           fired cator         0.003251288448185896           on         0.003251288448185896           on         0.003251288448185896           on         0.003251288448185746           on         0.00325128846185896           on         0.002117200777766781           on         0.0011172007777667839048           on         0.00211	0.019168485084734677	OverTime No
Degree indicator 0.017973703576334765 0.017973703576334765 0.011686886688498488 0.011686886688498488 0.011686886688498488 0.010231729598750271 0.010231729598750271 0.010231729598750271 0.010231729598750271 0.010231729598750271 0.000993728410945356 0.00993728410945356 0.00993728410945356 0.00993728410945356 0.00993728410945356 0.00993728410945356 0.00993728410945356 0.00993728410945356 0.0099372841094335 0.0099372841094335 0.0099372841094335 0.0099372841094335 0.0099372841094335 0.0099372841094335 0.00993728441887246 0.009280447887246 0.0003217289441887246 0.0003217289441887246 0.0003217289441887246 0.000117200777765781 0.000117200777765781 0.000117200777765781 0.000117200777766738 0.000117200777766738 0.0001830364728639048 0.0001830364728847285	0.018468798646649853	Over11me re
Degree indicator         0.011686886688498488           0.01047963224052392           0.01047963224052392           actor         0.01003532930183954           arely indicator         0.009949488392666073           Development indicator         0.00993728410945356           Development indicator         0.00858154936167823           inector indicator         0.00858154936167823           indicator         0.0085817477189906           ond/82873777189906           ndicator         0.004812104734335           equently indicator         0.00481210473645395           for         0.00481210475045395           for         0.00428917471423435           equently indicator         0.004289174714234355           equently indicator         0.004889174124335           for         0.0042891271243           firstor         0.003294189881221243           on         0.002809269441787746           on         0.003251288441857746           0.001880646647395356           0.0018806466473953576           0.0018806466473953576           0.001880646647893812121243           0.001880646647893831873           0.00188064728639048           0.001880646647893831873           0.0	0.017973703576334765	DercentSalan
0.010447963224052392   2.0010447963224052392   2.00100953293129589750271   2.001009532930183954   2.001009532930183954   2.001009532930183954   2.001009532930183954   2.0010095329301839566073   2.001009949488392666073   2.001009949488392666073   2.001009949488392666073   2.0010999895295   2.001098995913108   2.001098995913108   2.001098995913108   2.001098995913108   2.001098995913108   2.001098995913108   2.00109899591231243   2.00109899594118827246   2.00109899594118827246   2.00117200777766781   2.0011989642953526   2.00117200777766781   2.00119896442953528   2.00118980546242953528   2.00118980546242953528   2.00118980546242953528   2.00118980546242953528   2.00118980546242953528   2.00118980546242953528   2.00118980546542953528   2.0011898054654295358   2.0011898958971857   2.0011898958971857   2.0011898958971857   2.0011898958971857   2.0011898958971857   2.0011898958971857   2.0011898958971857   2.00118989589787   2.00118989589787   2.00118989589787   2.00118989589787   2.00118989589787   2.00118989787   2.00118989787   2.00118989787   2.00118989787   2.001189898787   2.00118989787   2.00118989787   2.00118989787   2.00118989787   2.001189898787   2.001189898787   2.001189898787   2.00118989787   2.001189898787   2.00118989787   2.001189898787   2.001189898787   2.00118989787   2.00118989787   2.001189898787   2.001189898787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988787   2.00118988	0.011686886688498488	YearsWithCur
active indicator 0.010231729589750271 active indicator 0.010094488392666073 active indicator 0.00099532930183954 active indicator 0.000993728410945356 bevelopment indicator 0.00993728410945356 bevelopment indicator 0.00838114274798098 indicator 0.00838114274798098 bridicator 0.007852873777189906 occupantly indicator 0.007852873777189906 occupantly indicator 0.007852873777189906 occupantly indicator 0.0048127104750485395 for 0.0048127104750485395 indicator 0.004289646706544303 fixture indicator 0.0032041898127143 firstor 0.003251288448185896 occupantly indicator 0.003251288448185896 occupantly indicator 0.003261288448185896 occupantly indicator 0.00326128848185896 occupantly indicator 0.003261288848185896 occupantly indicator 0.003261288848185896 occupantly indicator 0.003261288888888888888888888888888888888888	0.010447963224052392	StockOptionL
ratior 0.01009532930183954  rately indicator 0.009093728410945356  Development indicator 0.00993728410945356  Development indicator 0.008581548361678237  rector indicator 0.00838114274788098  j indicator 0.00785287377189906  ridicator 0.007852873777189906  ridicator 0.007852873777189906  ridicator 0.00440733758962592  rot 0.004812104750485395  rot 0.004812104750485395  rot 0.004812104750485395  rot 0.004812104750485395  rot 0.004812104750485395  rot 0.0048121043041887246  rot 0.003217200777766781  rot 0.001880646642953526  rot 0.0018980646642953526  rot 0.0018980646642953528  rot 0.0018980646642953528	0.010231729589750271	Environment
srely indicator         0.009949488392666073           Development indicator         0.009093728410945356           Development indicator         0.00858154936167823           Development indicator         0.008383114274798098           Indicator         0.008383114274798096           Indicator         0.0078287377189906           Indicator         0.0078287377189906           Indicator         0.007821714718234235           equently indicator         0.004289147174342355           tor         0.004812210475045395           tor         0.004289174174234235           tor         0.00428917417423435           incator         0.0042891212143           on 0.00212188441887246         0.003280418981221243           on 0.0021717807477766781         0.002809269441787746           on 0.002171200777766781         0.0019806466427953576           on 0.001830364728639048         0.001830364728639048           on 0.0018303665077766738           on 0.001830364728639048           on 0.001830364728639048           on 0.0018303665057766738	0.01009532930183954	YearsInCurrer
st indicator 000903728410945356  Development indicator 0.00638124365678237  Development indicator 0.0063811427478908  indicator 0.00744073758659513108  nolicator 0.0074407375865592  nolicator 0.00748071471434235  equently indicator 0.004880714714234235  tor 0.004880714714234235  tor 0.004880714714234235  inclicator 0.0042896470654303  inclicator 0.004380418821246  0.002171200777766781  0.002171200777766781  0.001188064642953526  0.001188064642953526  0.0011806650578766738	0.009949488392666073	JobSatisfactio
Development indicator         0.008581548361678237           rector indicator         0.008323114274798098           j indicator         0.008323142777189906           noticator         0.007852873777189906           noticator         0.007852873777189906           noticator         0.00488071471423423455           control         0.004812710475048395           tor         0.004812710475048395           tor         0.004289646706544303           tor         0.0042812713786888755           rindicator         0.003820418891271243           0.003820418891271243         0.003820418891271243           0.00217200777766781         0.0018080646647853526           0.00117300777766781         0.0011830364728639048           0.0018130364728639048         0.0018130364728639048           0.0018130364728639048         0.0018130364728639048	0.009093728410945356	JobLevel
Indicator	0.008581548361678237	YearsSinceLa
y indicator 0.008323669059913108 0.00782873777128906 0.00748073758962592 0.0048807147147342355 equently indicator 0.0048807147147342355 tor 0.0048807147147342355 tor 0.0048807147147342355 tor 0.0048807147343393 in indicator 0.00487137368385755 incator 0.0038204188981251243 in indicator 0.003251288448185896 0.002805569417827246 0.002117200777765781 0.00113036472853526 0.00113036472853526 0.001130364728535378 0.00113036472853578	0.008383114274798098	TrainingTime
ndicator 0.007852813777199906  ndicator 0.00740733759962592  0.0048807147147342355  0.0048807147147342355  tor 0.004880714712342355  tor 0.00458122104750485395  ricator 0.004528964706544303  ricator 0.00320418981271243  ricator 0.00320418981271243  0.00217720077776578  0.00217720077776578  0.00189806466472853526  0.00189806466472853526  0.0018130364728639048	0.008323669059913108	WorkLifeBala
ndicator 0.007440733758962592 0.004880747142342355 equently indicator 0.004289646706544303 tor 0.004289646706544303 title indicator 0.00428961212143 ilicator 0.003204188981221243 0.003217200777766781 0.002117200777766781 0.0018980646642953526 0.0018130364728639048 0.0018130364728639048	0.007852873777189906	Relationships
equently indicator 0.0048807147142342355  equently indicator 0.004428964706544303  title indicator 0.004428964706544303  indicator 0.0038204188981221243  r indicator 0.0038204188981221243  r indicator 0.00325128848185896  0.002177200777766781  0.00188064642953526  0.0018130364728639048  0.001836269276766738  0.001836269276766738  0.001836269276766738	0.007440733758962592	Jobinvolveme
equently indicator 0.0048122104750485395  tor 0.004288646706544303  0.004288646706544303  10.00405713736838575  10.00405713736838575  10.0057380418891121243  10.005880418891121243  10.0058804188185396  10.00288058941787746  10.0013880646647953576  10.0013880646647953576  10.0013880646647953576  10.001380646673953576  10.00138064667953576  10.001380646673953576  10.001380646673953576  10.00138064728639048  10.00138064728639048  10.00138064871877  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487187  10.0013806487  10.0013806488   10.0013806488   10.0013806488   10.0013806488   10.0013806488	0.0048807147142342355	Education
tor 0.004289646706544303  titive indicator 0.0042896418981271243  o.003820418891271243  o.003820418891271243  o.003820418891271243  o.002171200777766781  o.00217200777766781  o.0018306466472853526  o.0018306466478639048	0.0048122104750485395	MaritalStatus
ative indicator 0.004057137368385755 irindicator 0.003251288448185896 o.002351288448185896 o.002351288448185896 o.002317209777766781 o.0018980646642953526 o.00181034678869048 o.00181036478659048 o.00181036478659048 o.0018102650276760738	0.004289646706544303	Businessirav
Iriator 0.0038204188981221243  r indicator 0.003251288448185896  0.0023092694417827246  0.00211720077766781  0.0019880646642953526  0.0018130364728639048  0.001836269275766738  0.00183626927566738	0.004057137368385755	John Cales
n indicator 0.0032112884481185896 0.0028092694417827246 0.00211720077766791 0.001388064664293525 0.0018130364728639048 0.001936265057766738 0.00193737268591787	0.0038204188981221243	John Sales
0.00280926941787246 0.002117200777766781 0.001980646642953526 0.0018130364728639048 0.00192650577566738	0.003251288448185896	Department
0.002117200777766781 0.0019806466429535.5 0.0018130364728639048 0.0015026050276760738 0.0015477774084831857	0.0028092694417827246	MaritalStatus
0.0019880646642953526 0.0012130364728639048 0.0015026050276760738 0.00124577174084871857	0.00211720077766781	Department F
0.0018130364728639048 0.0019026050276760738 0.00160260527174084871857	0.0019880646642953526	Gender Fema
0.0015026050276760738	0.0018130364728639048	MaritalStatus
0.00043677174084821857	0.0015026050276760738	BusinessTrav
1001001011100000	0.00043677174084821857	Gender Male
		0.04508039119933 0.033729208271344134 0.033729208271344134 0.02744582896673833 0.02169236715204148 0.02169236715204148 0.02169236715204148 0.01166886688498488 0.011058178668498488 0.010597865376334765 0.0105917895324052392 0.0105917895324052392 0.0105917895324052392 0.0099494883266073 0.0099494883266073 0.0098314274798998 0.00983144274798998 0.00983144274798998 0.00983144274798998 0.0098807147142342355 0.00488071474244818596 0.0048807147826394555 0.0048807147854818596 0.003217280441827246 0.0032172804778578659048 0.00117200777766781 0.00198064642953526 0.00117200777766781

### Features that O032380560778504 O032380511553864 O0323805115735609 O0323809112755009 O0323809112755009 Attrition:

Age

0.022292969362632458

0.020091790563335016 0.019420213693461752 0.01604789830137631

0.01580889690719255

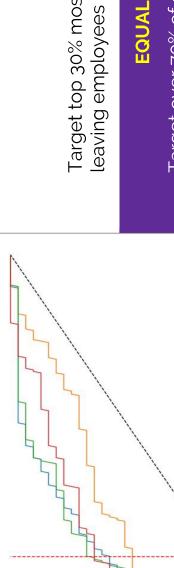
0.025785189748205522

- Income
- Working Overtime
  - Working Years
    Job Satisfaction

0.007443012758332482

Divorced indicator

# Cumulative Gain Curve



Logistic Regression
 Decision Tree
 Support Vector Machine
 Random Forest
 Random

%08

100% -

Target top 30% most possible

#### EQUALS

Target over 70% of employees who actually would leave!

### Solve the Problem

Main Idea: change part of an employee's status to make the employee more likely to stay.

#### Approach

- Naive approach: change some features of a employee and predict the turn off probability again and check if the attrition rate drops.
- Interactive approach: design a questionnaire for the employees who are detected to be most likely to leave and change the employee's status
- can be changed: Monthly Income, Job Level, Years at Current Role... cannot be changed: Age, Total Working Years, Gender... can be changed gradually: Job Satisfaction... Features that

## THANKS FOR LISTENING!

