EMPLOYEE ATTRITION

IBM Dataset: Presentation

Note:
This presentation is in collaboration with 6 analysts. Authors can be provided by request..

Background & Objective

Why Attrition Matters?

REPLACEMENT COST

Direct Exit Cost
Recruiting
Training

EMPLOYEE MORALE

Lost of Productivity
Disengagement
Domino Effect

KNOWLEDGE LOST

Institutional Knowledge
External Relationships
Service Quality

"For someone making $40K a year, replacement cost is $20K - $30K in recruiting and training expenses."
-- *Society of Human Resource Management*

Causes of Attrition

- Unsatisfying compensation and benefits
- Lack of development opportunity
- Lack of work-life balance
- Lack of recognition
- Poor management
- Poor work conditions

We want to answer...

Why do employees leave?

What factors characterize employee attrition?

What can companies do to prevent losing employees?

Agenda

Background & Objective
Research
Questions to Answer

Descriptive Analysis
Overview of Data
Challenges

Data Preprocessing
Over Sampling
Attribute Selection

Modeling
Classifier Selection
Feature Importance

Insights & Recommendation
What Companies Can Do

Descriptive Analysis

Overview of Data

IBM HR Employee
https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Structure

1470 observations

35 attributes

Label - Attrition

“Yes”: 237 (16%)

“No”: 1233 (84%)

Data Types

Numeric

Categorical

Ordinal/Scale

Challenges

Biased Dataset

The numbers of "Yes" and "No" are unbalanced

237 Yes
1233 No

Accuracy vs. Precision

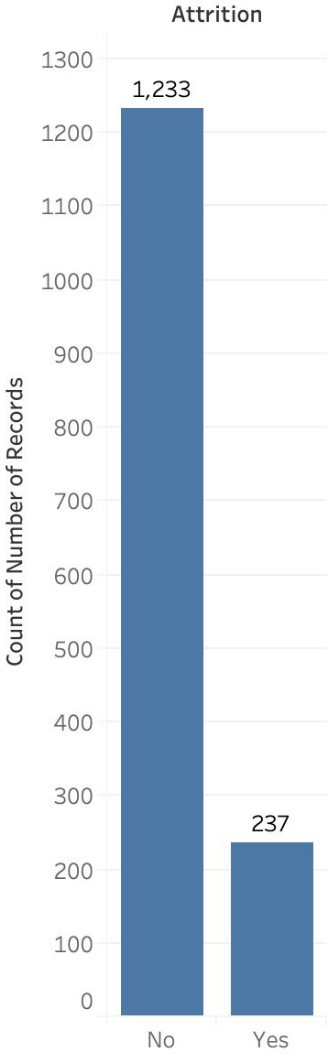
Need to focus on the number of 'Yes', instead of 'No'

TP/(TP+FP)

Too Many Attributes

Problem with overfitting and redundancy

35 attributes



Data Preprocessing

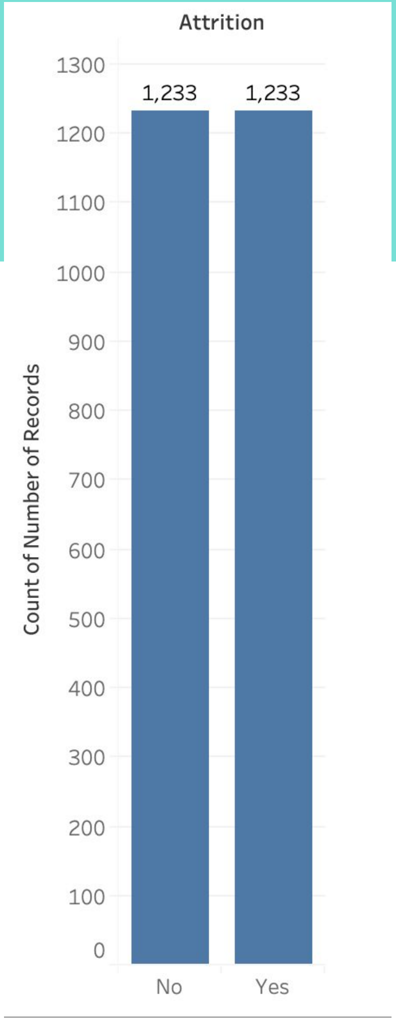
Remove Single Unique Value

- Employee ID
- Employee Count
- Over 18
- Standard Hours

Remove Highly Correlated Variables

	d.TotalWorkingYears	d.YearsAtCompany	d.YearsWithCurrManager	d.YearsInCurrentRole	d.YearsSinceLastPromotion
d.TotalWorkingYears	1.0000000	0.6281332	0.4591884	0.4603646	0.4048578
d.YearsAtCompany	0.6281332	1.0000000	0.7692124	0.7587537	0.6184089
d.YearsWithCurrManager	0.4591884	0.7692124	1.0000000	0.7143648	0.5102236
d.YearsInCurrentRole	0.4603646	0.7587537	0.7143648	1.0000000	0.5480562
d.YearsSinceLastPromotion	0.4048578	0.6184089	0.5102236	0.5480562	1.0000000

	d.MonthlyIncome	d.JobLevel	d.TotalWorkingYears
d.MonthlyIncome	1.0000000	0.9502999	0.7728932
d.JobLevel	0.9502999	1.0000000	0.7822078
d.TotalWorkingYears	0.7728932	0.7822078	1.0000000



Over Sampling | The“Yes”

Feature Selection

Top Features:

- Monthly Income
- Over Time
- Stock Option Level
- Years At Company
- Age
- Distance From Home

	Specs	Score
2	MonthlyIncome	411536.225257
5	YearsAtCompany	433.389238
0	Age	306.601455
1	DistanceFromHome	168.847410
13	OverTime	145.667368
14	StockOptionLevel	90.301831
12	MaritalStatus	52.232840
19	low_worklife_balance	31.705882
18	low_job_involvement	31.053763
17	frequent_travel	30.952381
9	EnvironmentSatisfaction	22.861395
3	NumCompaniesWorked	20.701826
11	JobSatisfaction	17.644864
16	low_relationship_satisfaction	10.695312
4	TrainingTimesLastYear	6.383004
6	Department	4.292568
10	Gender	1.063830
8	EducationField	0.805780
7	Education	0.453708
15	low_percentage_hike	0.326425

Modeling

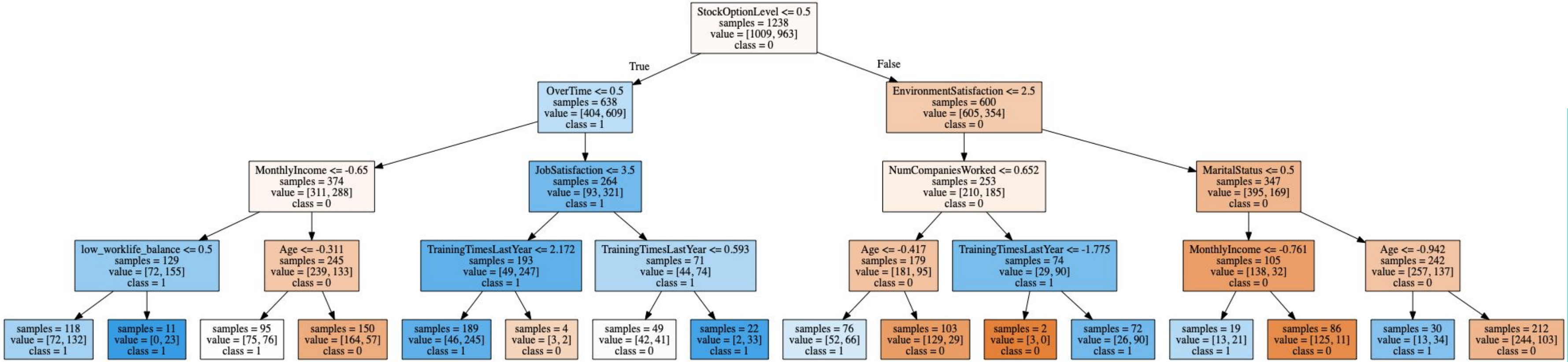
Modeling with All Attributes

Model	Accuracy	Precision
Logistic Regression	74.2%	73%
Decision Tree	78.9%	73.7%
Random Forest	78.1%	79.1%
Gradient Boosting	95.9%	92.4%

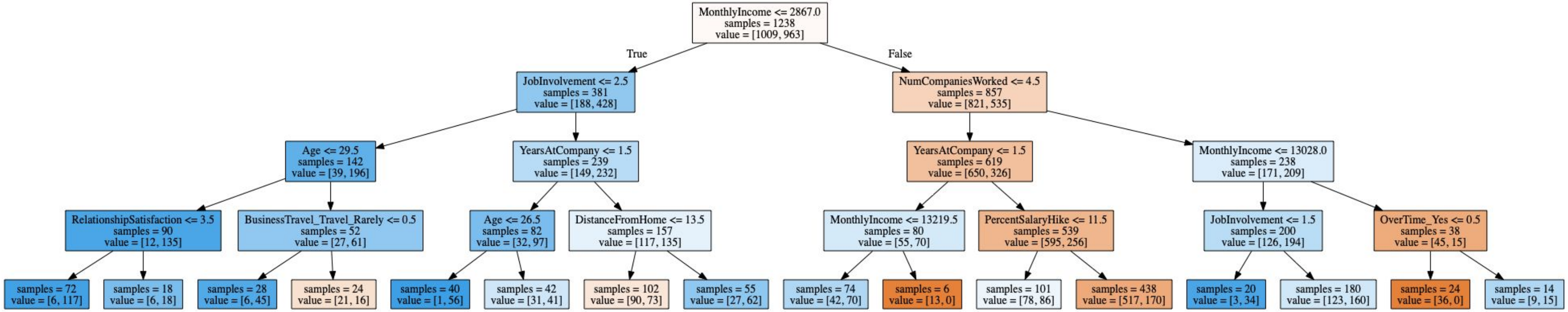
Modeling with Top Attributes from Feature Selection

Model	Accuracy	Precision
Logistic Regression	69.4%	67%
Decision Tree	75.5%	78.3%
Random Forest	75.9%	77.7%
Gradient Boosting	92.9%	87.4%

Decision Tree (Decision Nodes)



Random Forest (Decision Nodes)



Insights & Recommendation

Why do employees leave?

Theory of Organizational Equilibrium

An Employee will stay with an organization:

- If attributes such:
  - Satisfactory **Pay**
  - Working **Conditions**
  - Developmental **Opportunities**
- Are equal to or greater than:
  - Time / Effort

What we found influences Employees to Leave

Overtime

Time/effort

Monthly Salary

Satisfactory Pay

Job Involvement

Development

Age

Development

Stock Options

Satisfactory Pay

Years With Company

Working Conditions

Most Important Attributes

- Overtime
- Age
- Monthly Income
- Years At Company
- Stock Options

How to address?

- Training/Skills & Promotions
- Manage Age 26-34
- Promotions Opps. For Income below $2960
- First few years are Highest Risk
- Offer higher stock options

Characteristics of Attrition by Department

Insight 1

Employees with **technical degrees** are more likely to leave when working for **HR Department.**

85%

Insight 2

Employees from **all departments** are roughly twice as likely to leave when **working overtime**.

2X

Insight 3

Employees from all departments benefit from **High Job Involvement.**

73% /37%

Future of Employee Management

Employees that show signs of leaving...will **not only** be dealt with by managers and HR.... but by **solutions groups**, something IBM is already using today.

IBM has saved nearly **$300 million in retention costs** using similar AI and predictive techniques.

Source: https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html


THANKS
Q&A

Appendix

Initial Modeling

Our Best Models:

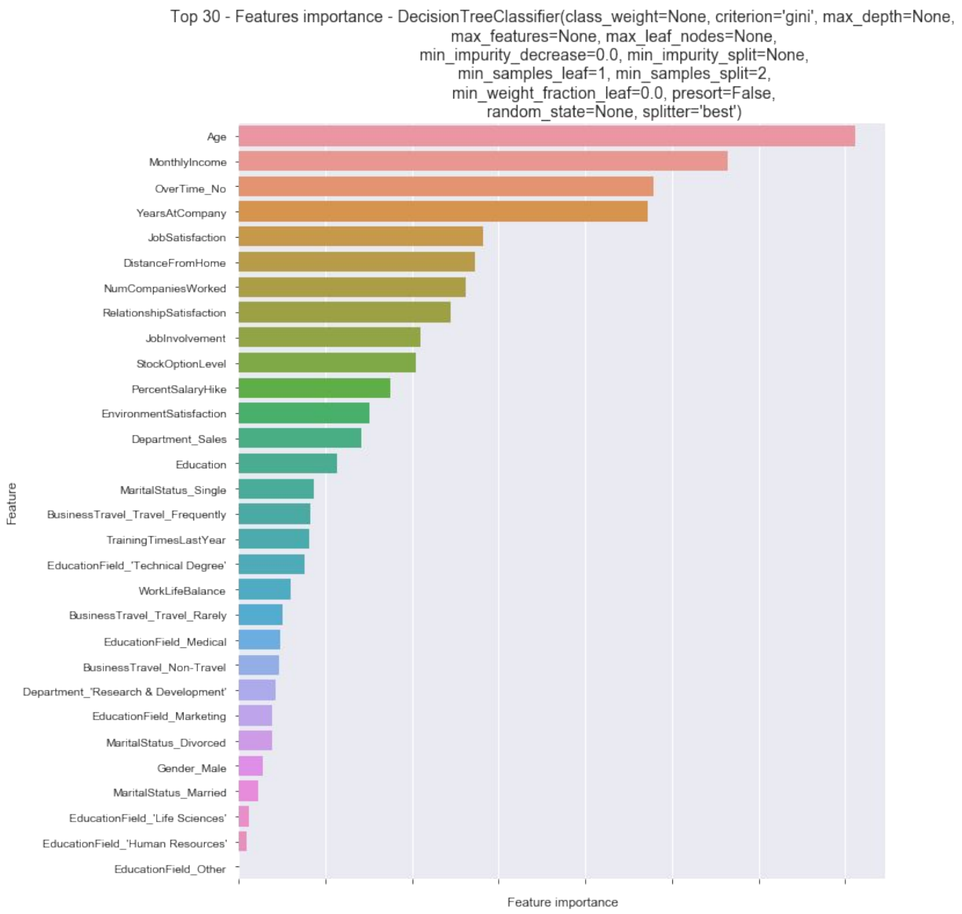
1. Support Vector Machine
2. Random Forest
3. Decision Tree
4. Gradient Boosting

	Classifiers	Crossval Mean Scores
0	Logistic Reg.	0.743309
1	SVC	0.982157
2	KNN	0.744120
3	Dec Tree	0.912003
4	Grad B CLF	0.874696
5	Rand FC	0.964315
6	Neural Classifier	0.572182
7	Naives Bayes	0.572182

Top Features
Decision Tree

Feature Importance:
Based on Gini Index

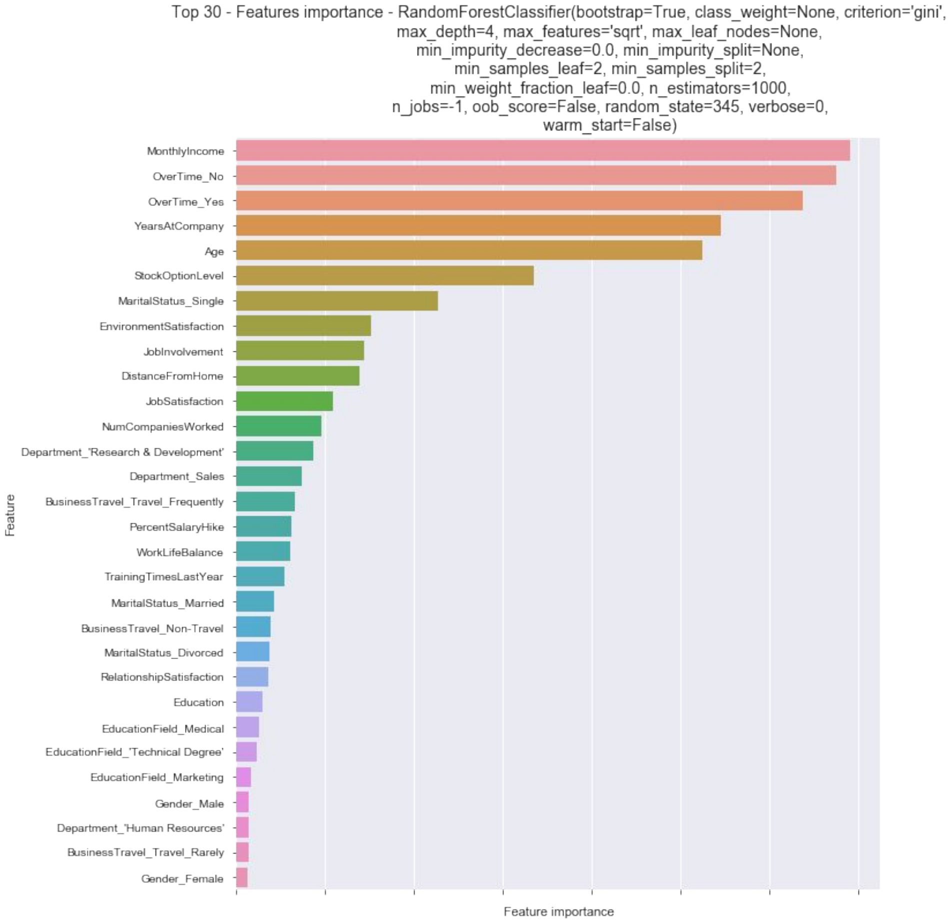
Top Five
1. Age
2. Monthly Income
3. Overtime_No
4. Years at Company
5. Job Satisfaction



Top Features
Random Forest

Feature Importance
Based on Gini Index

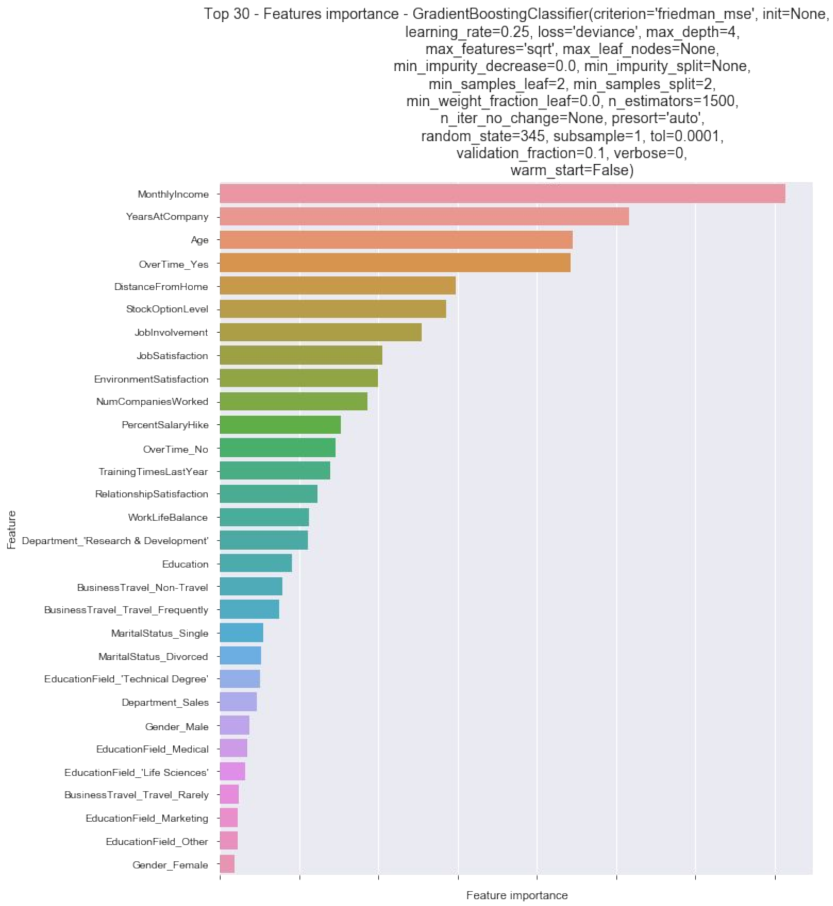
Top Five
1. Monthly Income
2. Overtime_No
3. Overtime_Yes
4. Years at Company
5. Age



Top Features
Gradient Boosting

Feature Importance:
Based on Friedman

Top Five
1. Monthly Income
2. Years At Company
3. Age
4. Overtime_Yes
5. Distance From Home



Highest Attrition Ratio

Attrition	No	Yes	All	YesToNo	YesOverTot
TotalWorkingYears					
40	0	2	2	inf	100.00
1	41	40	81	97.56	49.38
0	6	5	11	83.33	45.45
2	22	9	31	40.91	29.03
37	4	0	4	0.00	0.00
38	1	0	1	0.00	0.00

The highest ratio of attrition is in the first three years with the company. Between 30%-50% attrition.

Attrition	No	Yes	All	YesRatNo	YesRatio
Age					
19	3	6	9	200.00	66.67
20	5	6	11	120.00	54.55
18	4	4	8	100.00	50.00
58	9	5	14	55.56	35.71
59	10	0	10	0.00	0.00
60	5	0	5	0.00	0.00

The highest turnover rate is between the ages of 18-20 with an average turnover of 57%. Ages 59-60 saw no turnover, while 58 saw a turnover of 35%.

Largest Disparity

The greatest disparity in turnout is within Job Role, Age, and Job Level.

Attrition	No	Yes	All	YesRatio	YesRatNo
JobLevel					
1	400	143	543	26.34	35.75
2	482	52	534	9.74	10.79
5	64	5	69	7.25	7.81
4	101	5	106	4.72	4.95

For Job Role there is difference of up to 6x between "Sales Representative" and "Healthcare Representative".

Attrition	No	Yes	All	YesToNo	YesOverTot
JobRole					
Sales Representative	50	33	83	66.00	39.76
Laboratory Technician	197	62	259	31.47	23.94
Human Resources	40	12	52	30.00	23.08
Sales Executive	269	57	326	21.19	17.48
Research Scientist	245	47	292	19.18	16.10
Healthcare Representative	122	9	131	7.38	6.87

For Job Level, there is difference of up to 7x between "Level 1" and "Level 4".