ATTRITION STUDY

Improving employee retention using machine learning

JUNE 2020

Shivamm Gupta

EDINBURGH

cost of replacing a highly valued employee

up to **200%**

of their annual salary [1]

average expenditures are £30K [11]

MOTIVATION

- excessive replacement costs
  - role advertising costs
  - training costs
  - timely costs (adjustment period, ...)
- 87% of HR leaders consider improving retention a critical or high priority [1]
- 50% of all organizations globally struggle to retain their most valuable employees [1]

[1] https://blog.bonus.ly/10-surprising-employee-retention-statistics-you-need-to-know

ATTRITION V RETENTION

factors that contribute to attrition and retention based on importance [1-11]

- no opportunity for growth
- lack of recognition
- unsatisfied with their boss
- burn out
- lack of cultural fit
- job security
- happiness / satisfaction
- engagement with the role
- training allowing for growth
- feeling useful

[1] SOURCES HERE

USING DATA TO IMPROVE RETENTION RATES

IBM HR ANALYTICS EMPLOYEE ATTRITION & PERFORMANCE

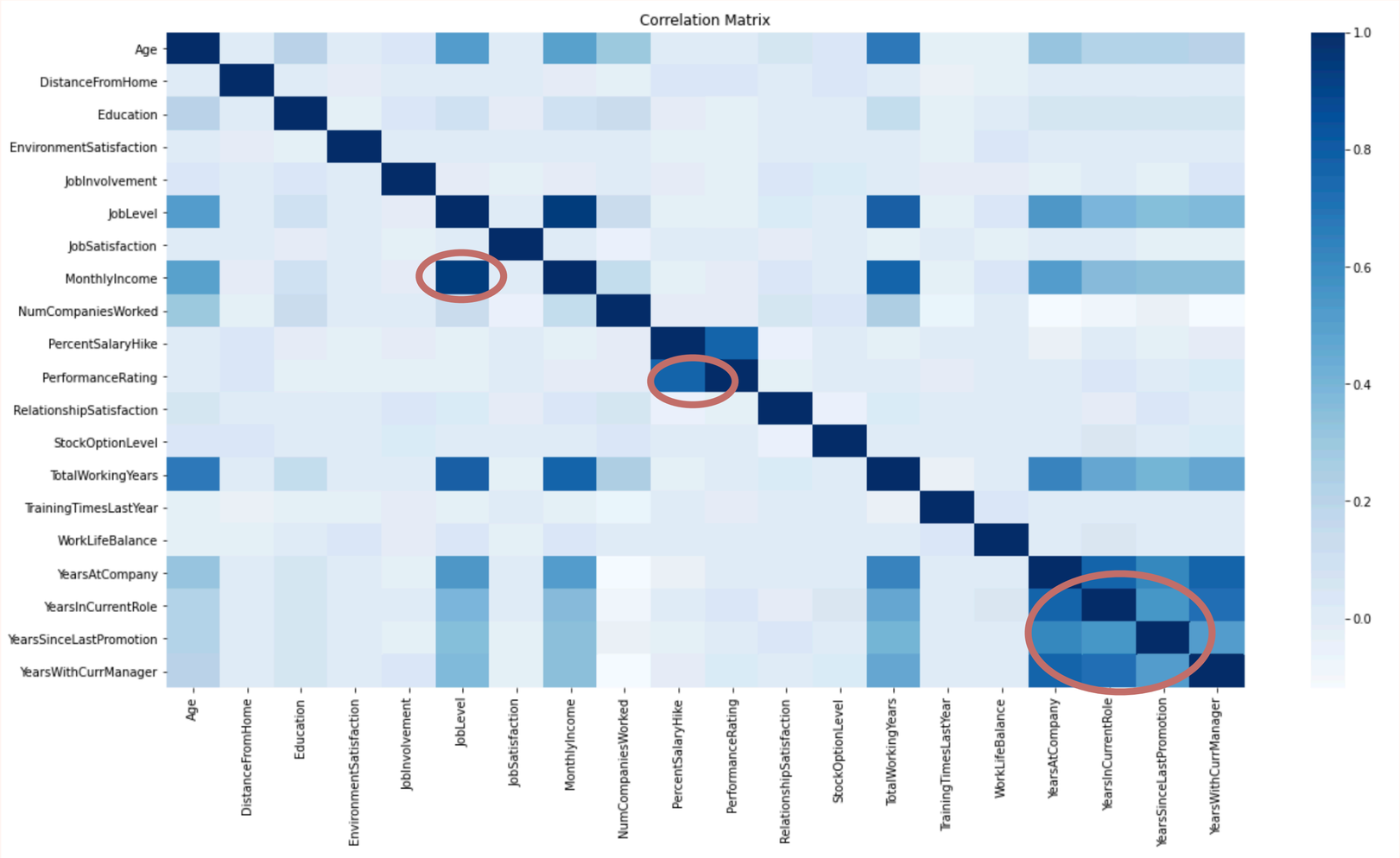
- fictional dataset created by IBM scientists [12]
- 1470 employee data points across 35 features
  - age, job level, monthly income, stock option, work-life balance,...

[12] https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

FEATURE CORRELATION

obvious correlations

- age
- job level
- monthly income / raise
- performance rating

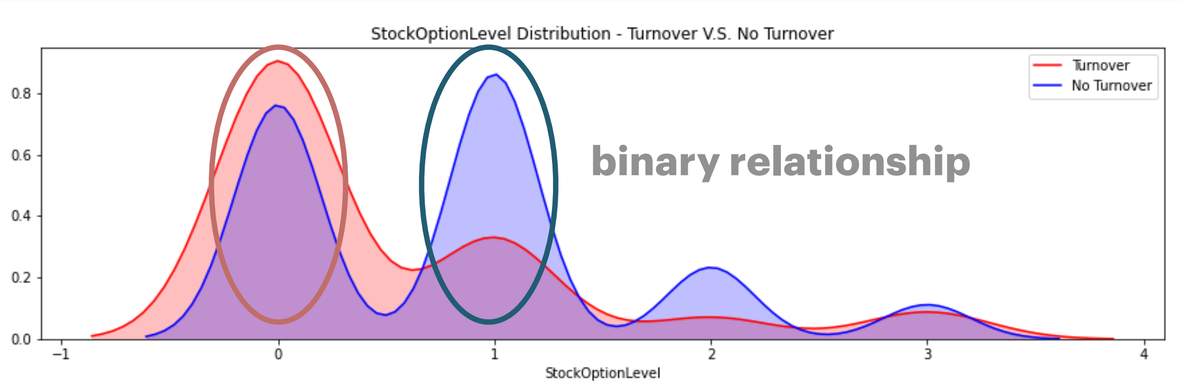


noteworthy observations

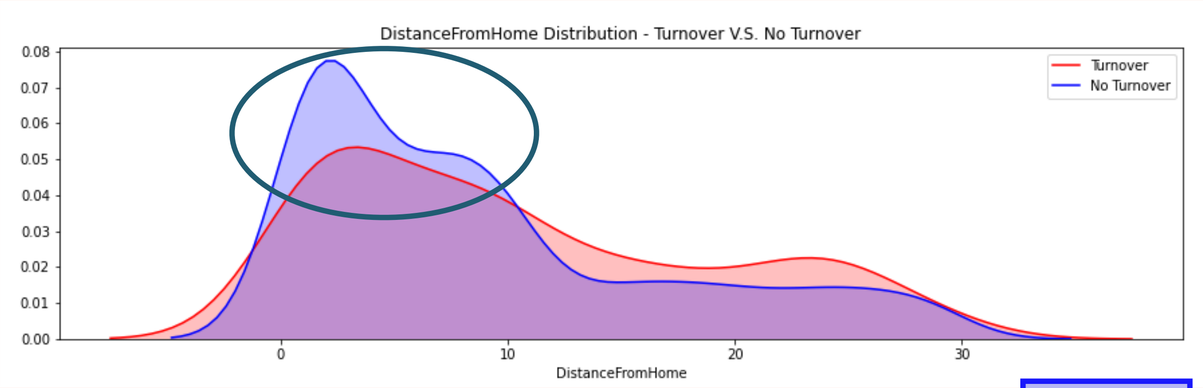
- years in current role - lack of promotion
- years in company - monthly income

} important retention factors?

NUMERICAL FEATURE ANALYSIS



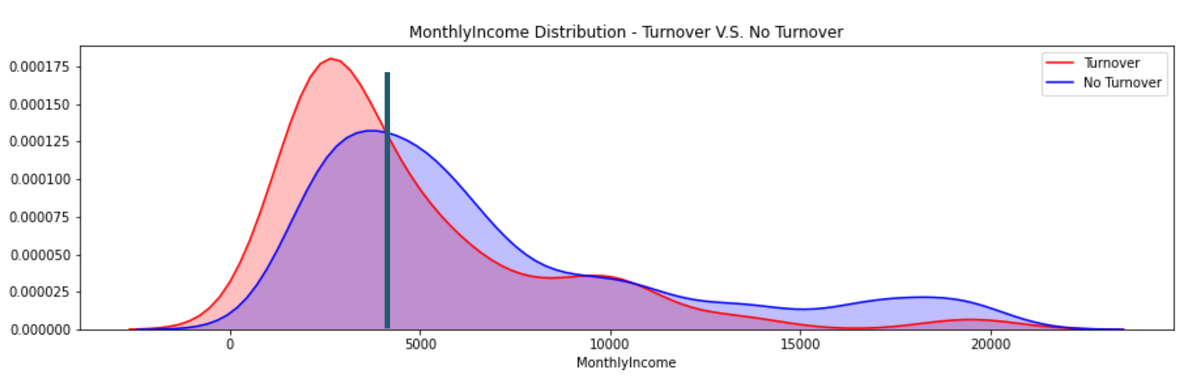
opportunity to buy stock



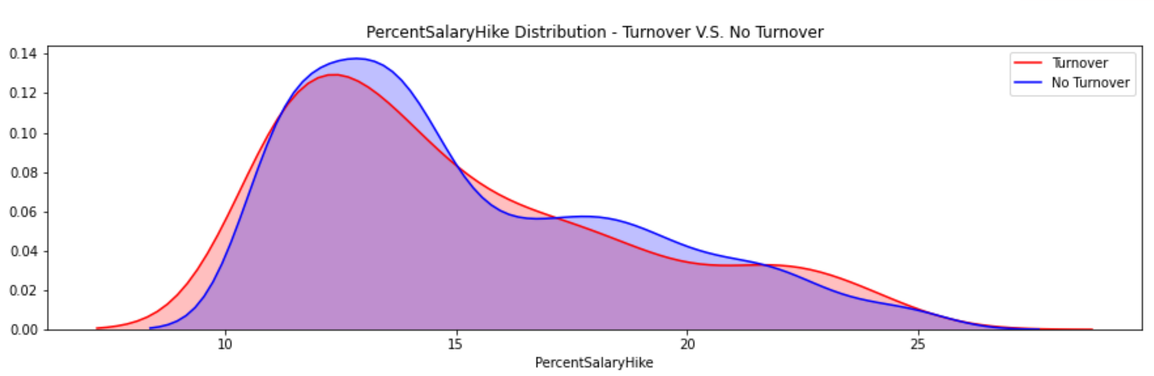
close commute

remain

leave

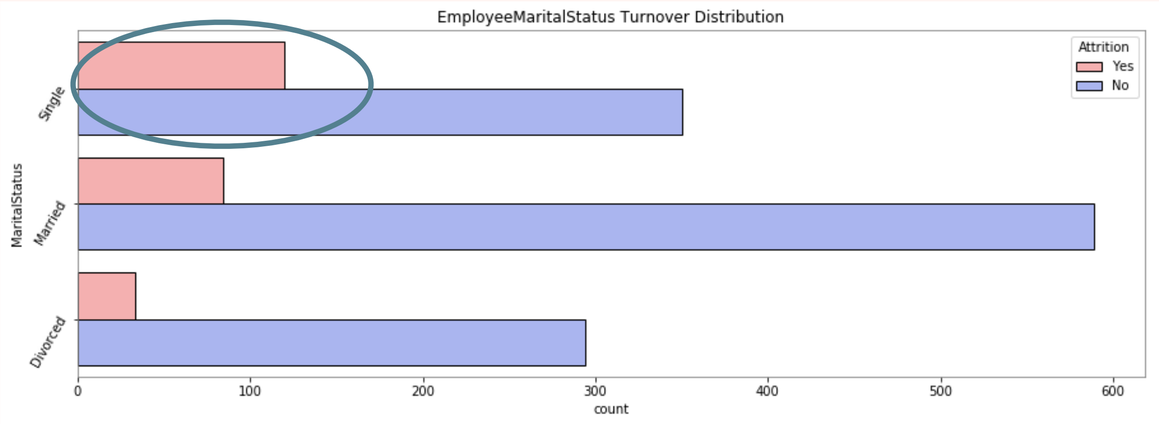


monthly income above threshold value

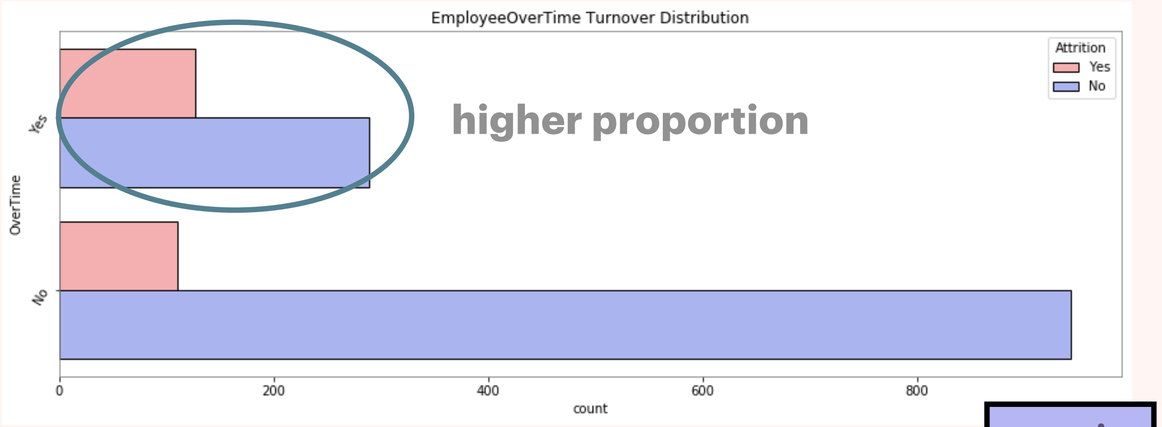


salary increase makes little difference

CATEGORICAL FEATURE ANALYSIS



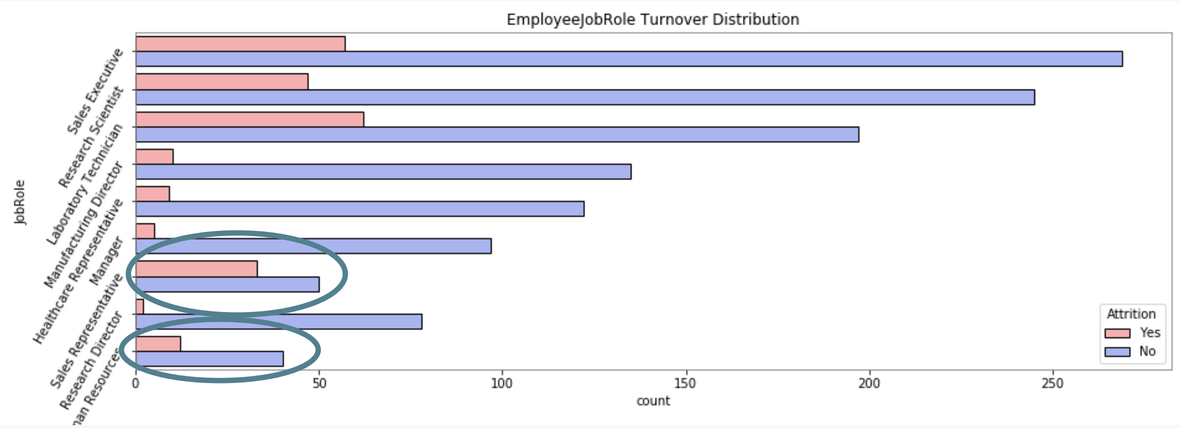
marital status - singles at higher risk



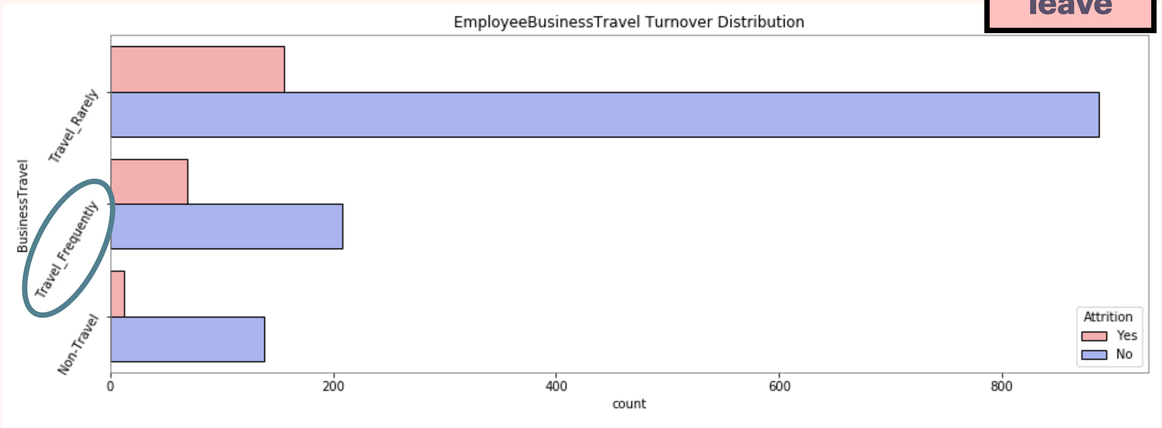
excess overtime

remain

leave



vulnerability: sales, HR



excess business travel

MACHINE LEARNING

NEURAL NETWORK

SUPPORT-VECTOR MACHINES

PRECISION

how many flagged employees left?

TP / (TP + FP)

ACCURACY

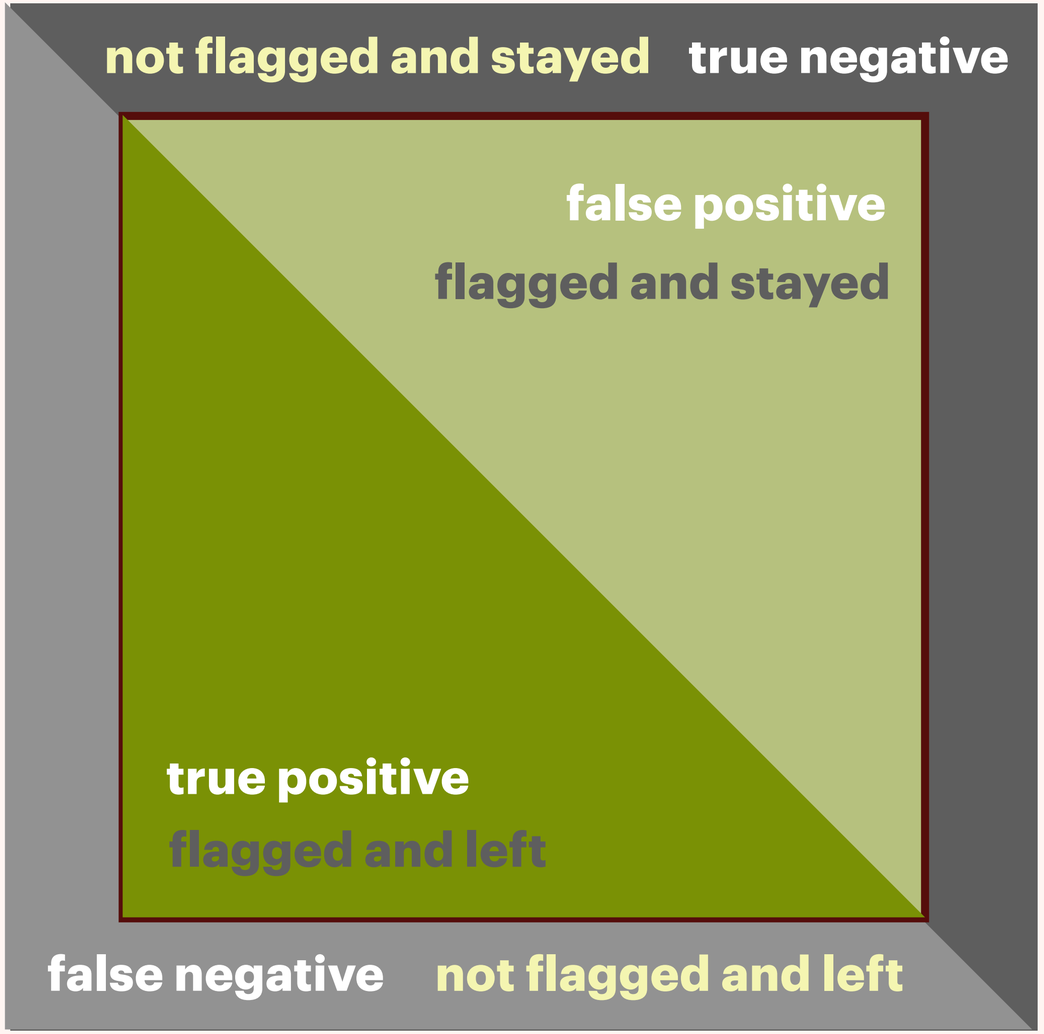
how many were correctly predicted in total

(TP + TN) / Total

RECALL

how many leaving employees were flagged?

TP / (TP + FN)



USING

PRECISION

how many flagged employees left?

TP / (TP + FP)

maximizes number of people that are flagged correctly

important since employee replacement is costly

if not costly choose recall

decreases recall

PRECISION - RECALL TRADEOFF

ACCURACY

how many were correctly predicted in total

(TP + TN) / Total

RECALL

how many leaving employees were flagged?

TP / (TP + FN)

reasonably high accuracy if no employees flagged - fails to fulfill purpose of model

MACHINE LEARNING

NEURAL NETWORK

non-explainable model

cannot identify features that lead to attrition

produces good results in general

SUPPORT-VECTOR MACHINES

RESULTS

- **easily updated and maintained - long-term advantage**
- **requires follow-up discussion with employee to identify features**
  - **leads to higher costs to prevent attrition**

train

	precision	recall	f1-score	support
0	0.89	1.00	0.94	863
1	0.95	0.37	0.53	166
accuracy			0.90	1029
macro avg	0.92	0.68	0.74	1029
weighted avg	0.90	0.90	0.87	1029

test

		precision	recall	f1-score	support
remain	0	0.88	0.99	0.94	370
leave	1	0.88	0.32	0.47	71
accuracy				0.88	441
macro avg		0.88	0.66	0.70	441
weighted avg		0.88	0.88	0.86	441

MACHINE LEARNING

NEURAL NETWORK

SUPPORT-VECTOR MACHINES

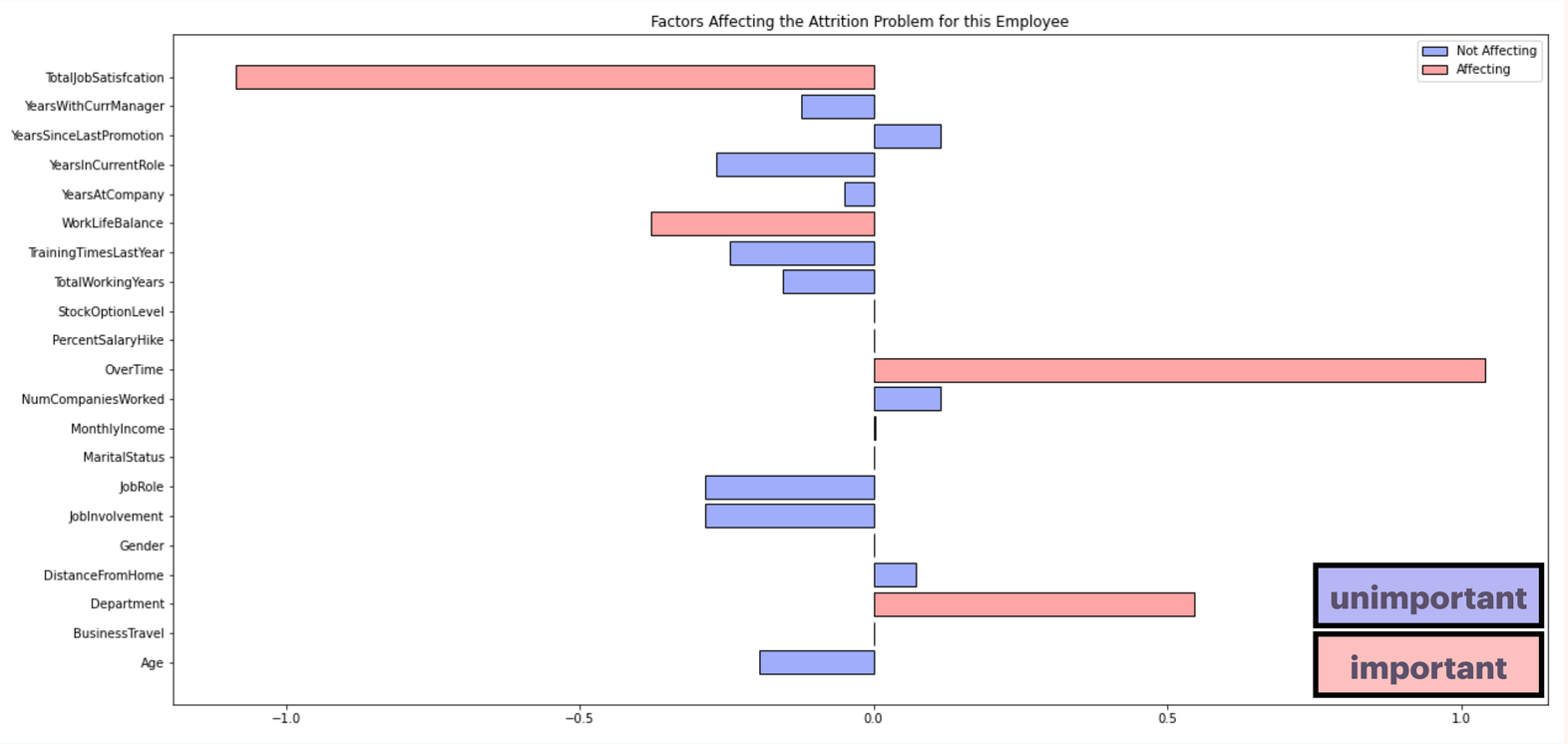
explainable model

identifies features that lead to attrition

more challenging to maintain good results

EXPLAINABILITY

- flags deciding features for each employee
- retention attempts easily implemented
- not achievable in neural network model



given employee feature breakdown

RESULTS

- produces better results at the moment
- requires a lot of data storage and retraining in the future - hard maintainability
- lower cost of attrition prevention if model trained well

train

	precision	recall	f1-score	support
0	0.90	0.99	0.94	863
1	0.88	0.40	0.55	166
accuracy			0.89	1029
macro avg	0.89	0.69	0.74	1029
weighted avg	0.89	0.89	0.88	1029

test

		precision	recall	f1-score	support
remain	0	0.90	0.98	0.94	370
leave	1	0.83	0.42	0.56	71
accuracy				0.89	441
macro avg		0.87	0.70	0.75	441
weighted avg		0.89	0.89	0.88	441

441

employees in test dataset

71

left

Could we have stopped them?

441

employees in test dataset

71

left

SVM Model flagged

36

employees

out of which

30

left

at least

£900K

SAVED if flagged employees convinced to stay [11]

in addition to time and effort needed to train new employees

precision: 83%

recall: 42%

MACHINE LEARNING

NEURAL NETWORK

- use for long-term purposes
- can be constantly updated

SUPPORT-VECTOR MACHINES

- simpler to decrease attrition in practice
- cheaper since follow-up meetings not necessary
- better for short-term purposes

MACHINE LEARNING CAVEATS

the 100% precision model

- all flagged employees want to leave
- hugely decreases recall
  - fails to flag many leaving employees
- fails to help HR decrease attrition effectively

train

	precision	recall	f1-score	support
0	0.84	1.00	0.91	863
1	1.00	0.01	0.01	166
accuracy			0.84	1029
macro avg	0.92	0.50	0.46	1029
weighted avg	0.87	0.84	0.77	1029

test

		precision	recall	f1-score	support
remain	0	0.84	1.00	0.91	370
leave	1	1.00	0.03	0.05	71
accuracy				0.84	441
macro avg		0.92	0.51	0.48	441
weighted avg		0.87	0.84	0.78	441

Logistic Regression

MACHINE LEARNING CAVEATS

why 71/71 is not possible

- imbalanced dataset
  - only 15-20% employees may want to leave
- high recall decreases confidence in flagged employees
  - will flag many employees that do not want to leave at all
- leads to wastage of HR resources
- cannot represent ALL factors that lead to attrition
  - better offer, family issues, metal health,...

THE MISSING PIECES

- more personal data - better recall and precision
  - happiness, training feedback, family status, work-life balance expectations,...
- assess employee demands and expectations
- **generic model that can be personalized is difficult**

IMPROVEMENTS

- feature engineering to make results better
- suggest specific ways to retain employee based on explainable model
  - predict income / raise required to retain employee

FIND OUT MORE

- GitHub
- Report
- Dataset

1 https://blog.bonus.ly/10-surprising-employee-retention-statistics-you-need-to-know
2 https://www.docebo.com/press/docebo-workplace-survey-report/
3 https://www.contactmonkey.com/blog/employee-engagement-trends
4 https://www.inc.com/todd-nordstrom/79-percent-of-employees-quit-because-theyre-not-ap.html
5 https://thriveglobal.com/stories/the-2019-rise-in-job-stress-and-burnout/
6 https://hbr.org/2019/12/burnout-is-about-your-workplace-not-your-people
7 https://www.forbes.com/sites/rachelmontanez/2019/06/05/burnout-is-sabotaging-employee-retention-three-things-you-must-know-to-help/#2413af135f0e
8 https://integrity-asia.com/blog/2018/11/21/80-employee-turnover-is-caused-by-bad-hiring-decision-here-are-the-5-costs-suffered-by-the-company/
9 https://daylightresources.co.uk/how-to-successfully-manage-a-large-team/
10 https://www.morganphilips.com/en/insights/articles/3-ways-to-incorporate-flexible-working-into-your-company-culture
11 https://www.hrreview.co.uk/hr-news/recruitment/it-costs-over-30k-to-replace-a-staff-member/50677