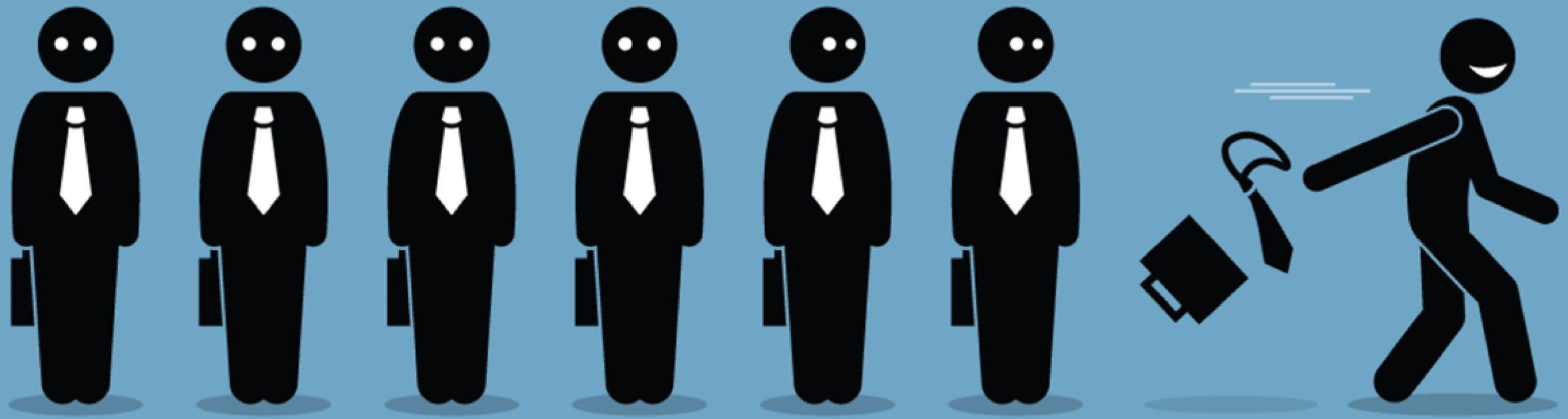
Employee churn predictions and survival analysis



Sona Khloyan, Pertchuhi Proshyan, Rafayel Mnatsakanyan

MGMT325 - Business Analytics

Introduction

Most typical reasons for turnover:

- Absence of opportunity for growth or career development
- Career Advancement
- Internal promotion
- Feeling overworked/burnout
- Opposing feelings towards supervisor or leadership
- Unhealthy work environment

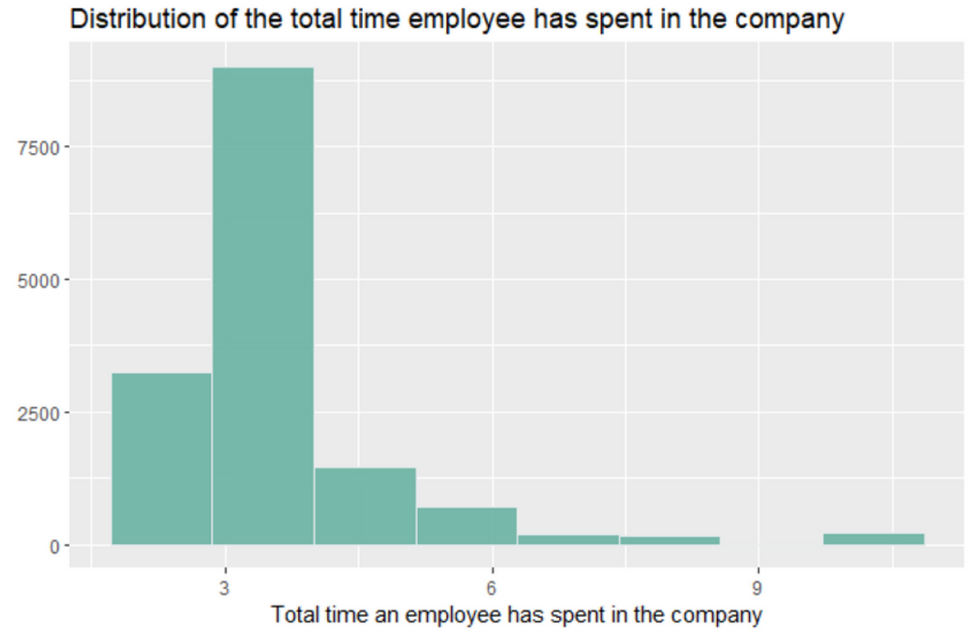
Objectives of the research

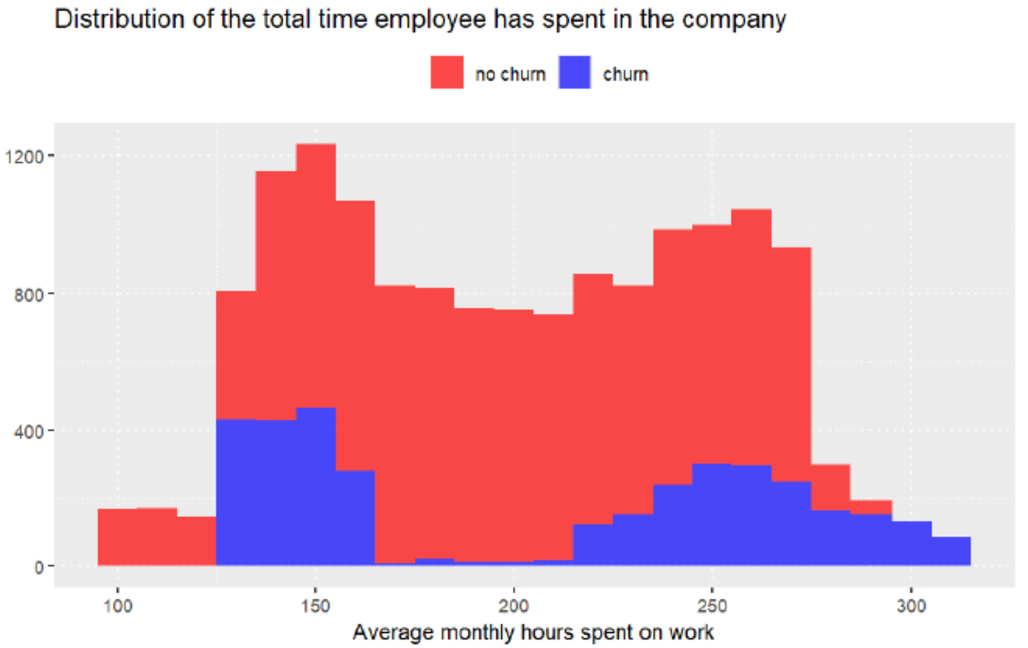
- to predict whether the employee will churn or not
- finding the factors that most influence the decision to churn
- to analyze the survival rates of employees in a company

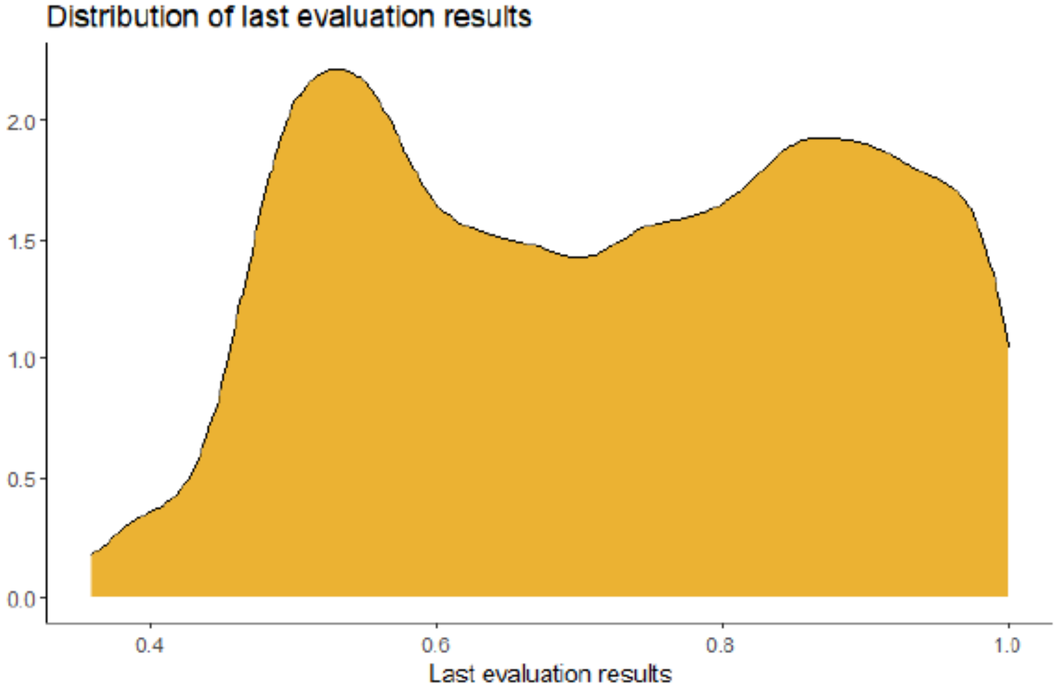
Data

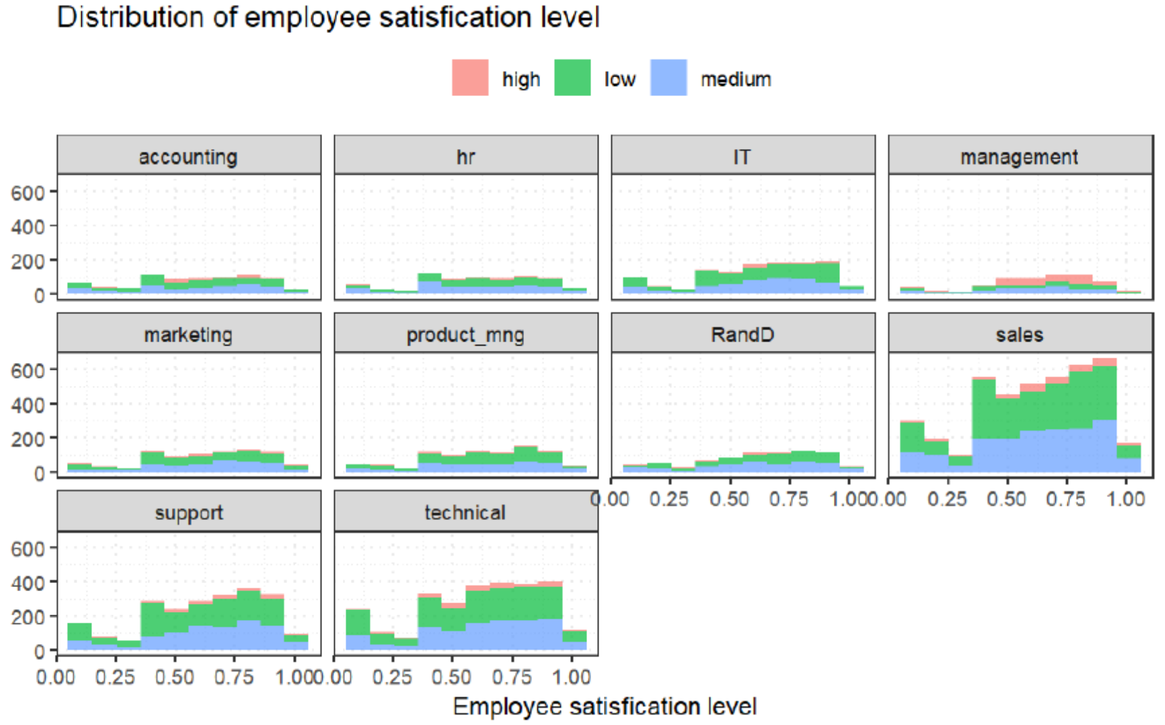
- **10 var**
- **15000 obs**

Variable name	Explanation
satisfaction_level	Employee satisfactory level.
last_evaluation	Results of the last performance evaluation conducted in the company.
number_project	Number of projects an employee conducts.
average_montly_hours	Average monthly hours an employee spends in the company.
time_spend_company	The total time an employee has spent in the company.
work_accident	Binary variable showing whether the employee has had a work accident in the company or not.
churn	Binary variable showing whether the employee will quit the company or stay.
promotion_last_5years	Binary variable showing whether the employee has got promoted during the last five years or not.
department	In which department the employee works.
salary	Categorical Salary: High, Medium or Low

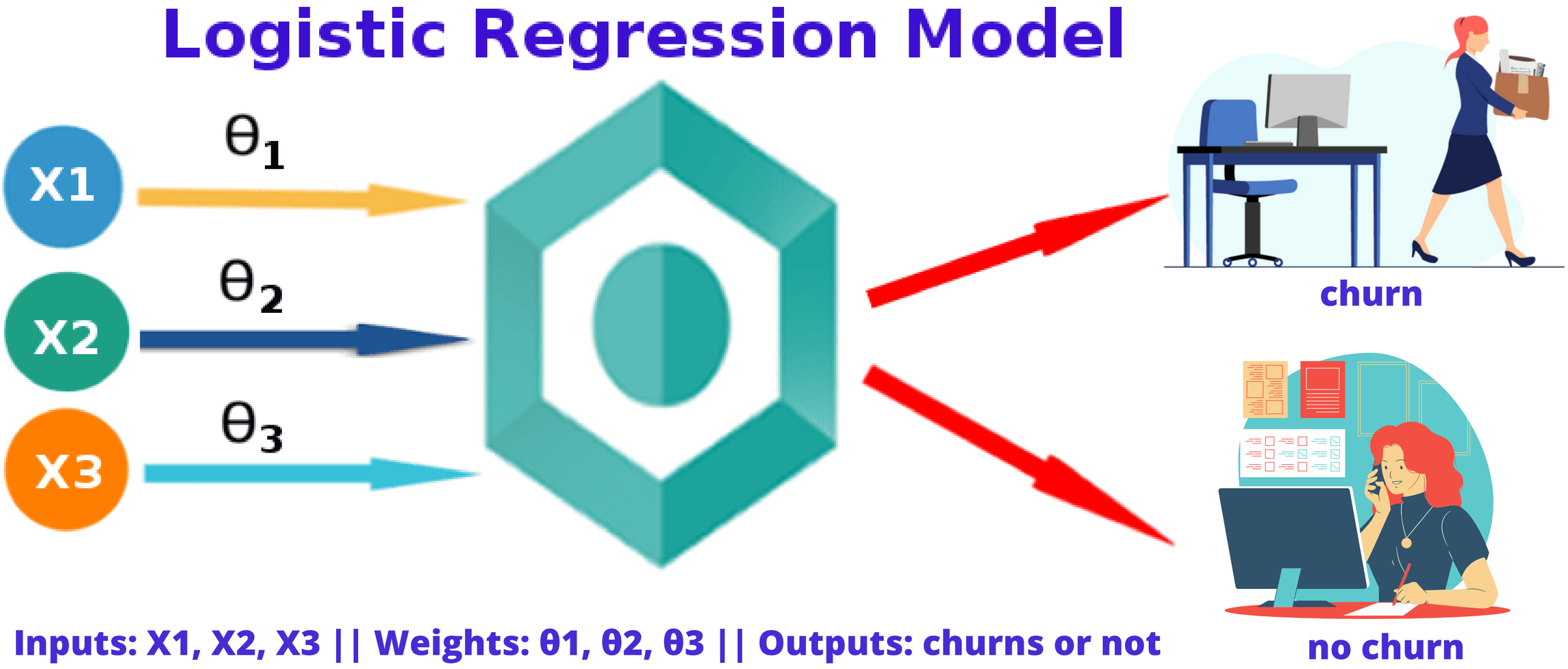

Distribution of the total time employee has spent in the company
7500
5000
2500
0
3
6
9
Total time an employee has spent in the company


Distribution of the total time employee has spent in the company
no churn
churn
1200
800
400
0
100
150
200
250
300
Average monthly hours spent on work


Distribution of last evaluation results
2.0
1.5
1.0
0.5
0.0
0.4
0.6
0.8
1.0
Last evaluation results


Distribution of employee satisfication level
high
low
medium
accounting
hr
IT
management
marketing
product_mng
RandD
sales
support
technical
600
400
200
0
0.00
0.25
0.50
0.75
1.00
Employee satisfication level

Methodology



Inputs: X1, X2, X3 || Weights: θ1, θ2, θ3 || Outputs: churns or not

Methodology: Logistic regression

Marginal Effects

	dy/dx	std err	z	P>|z|
salary[T.low]	0.3185	0.020	16.149	0.000
salary[T.medium]	0.2298	0.020	11.445	0.000
number_project	-0.0131	0.003	-4.399	0.000
time_spend_company	0.0435	0.002	19.995	0.000
work_accident	-0.2510	0.014	-18.286	0.000
promotion_last_5years	-0.2468	0.040	-6.105	0.000
average_montly_hours	0.0005	7.38e-05	7.070	0.000

- **Unit increase in years an employee has spent in a company is expected to increase the probability of churn by 4.35%.**
- **People getting low or medium salaries have more chances to churn the company than those with high salaries. The probability is increased by 23-32%.**
- **People who have got promotions during the last five years are 25% less likely to leave than those who haven't got one.**

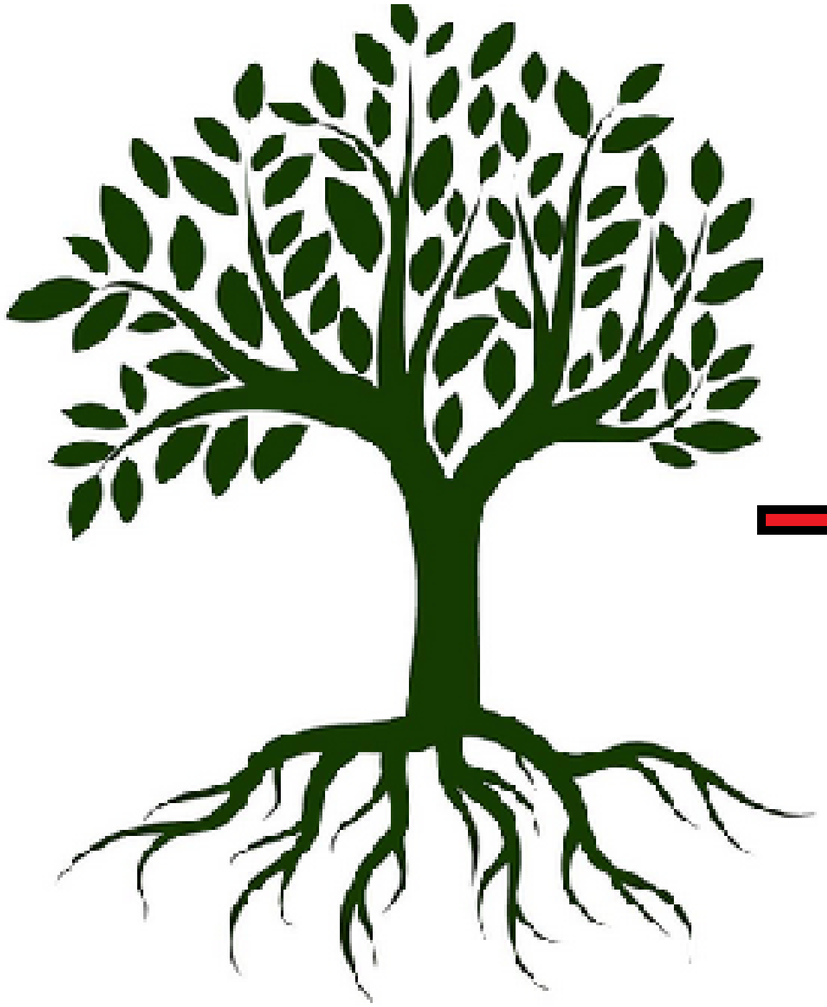
Methodology: Logistic regression

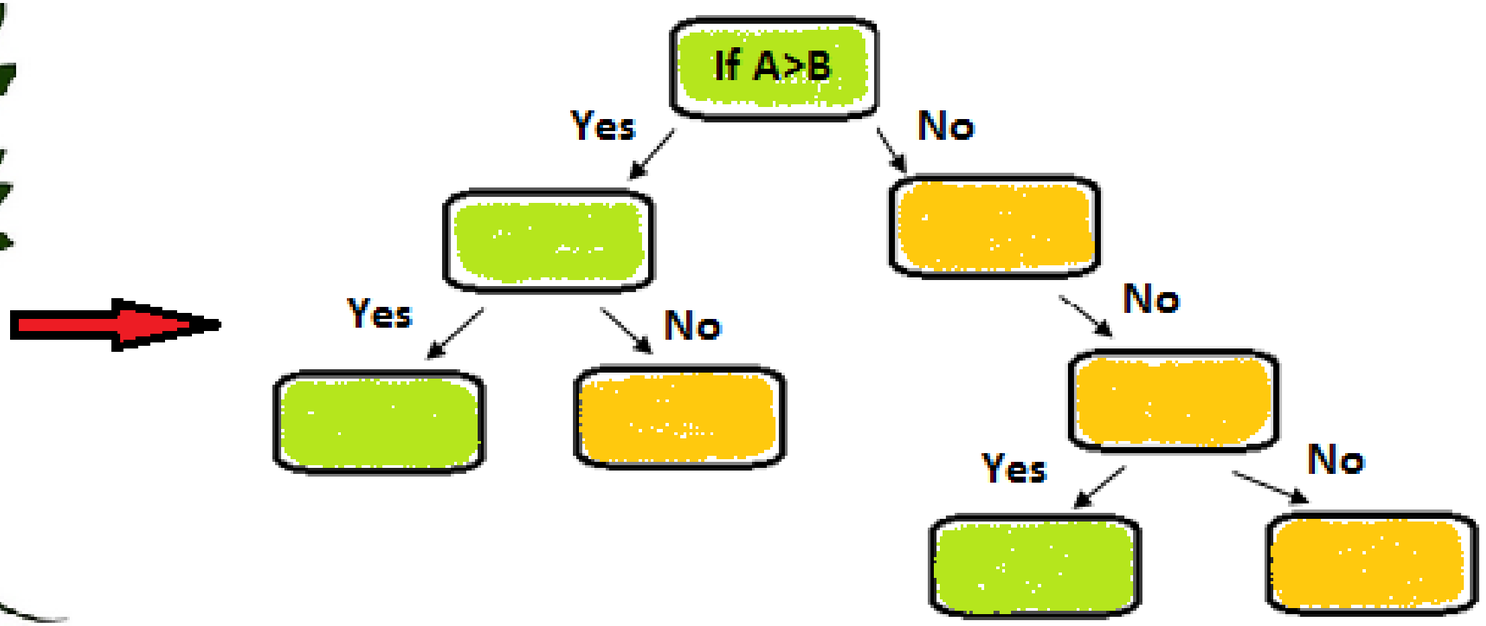
Conclusions

- **The amount of salary affects employees’ decision to churn the most: lower-paid employees tend to churn out the company.**
- **Also, each additional year spent in the company brings closer employee churn.**

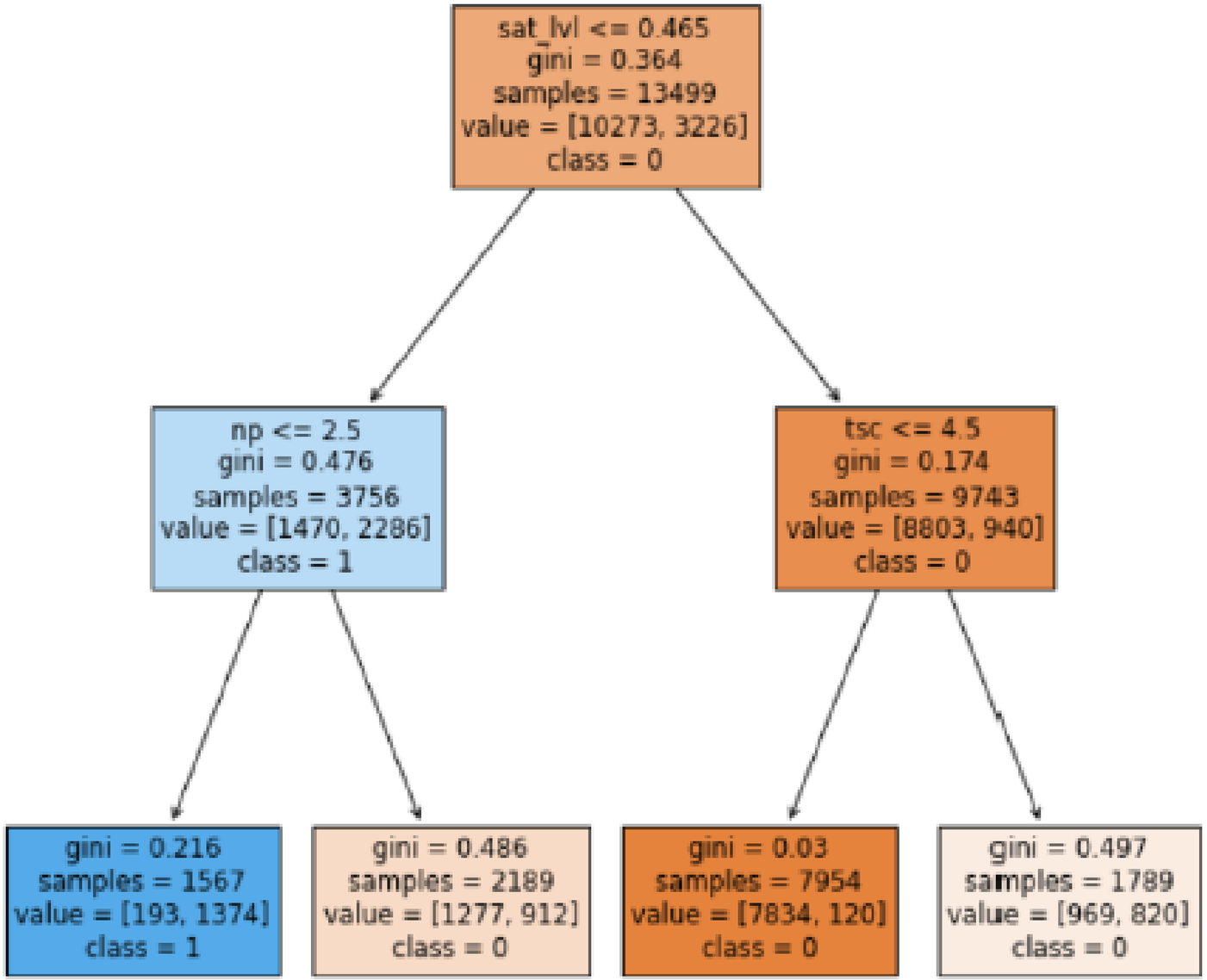
Methodology

Classification tree





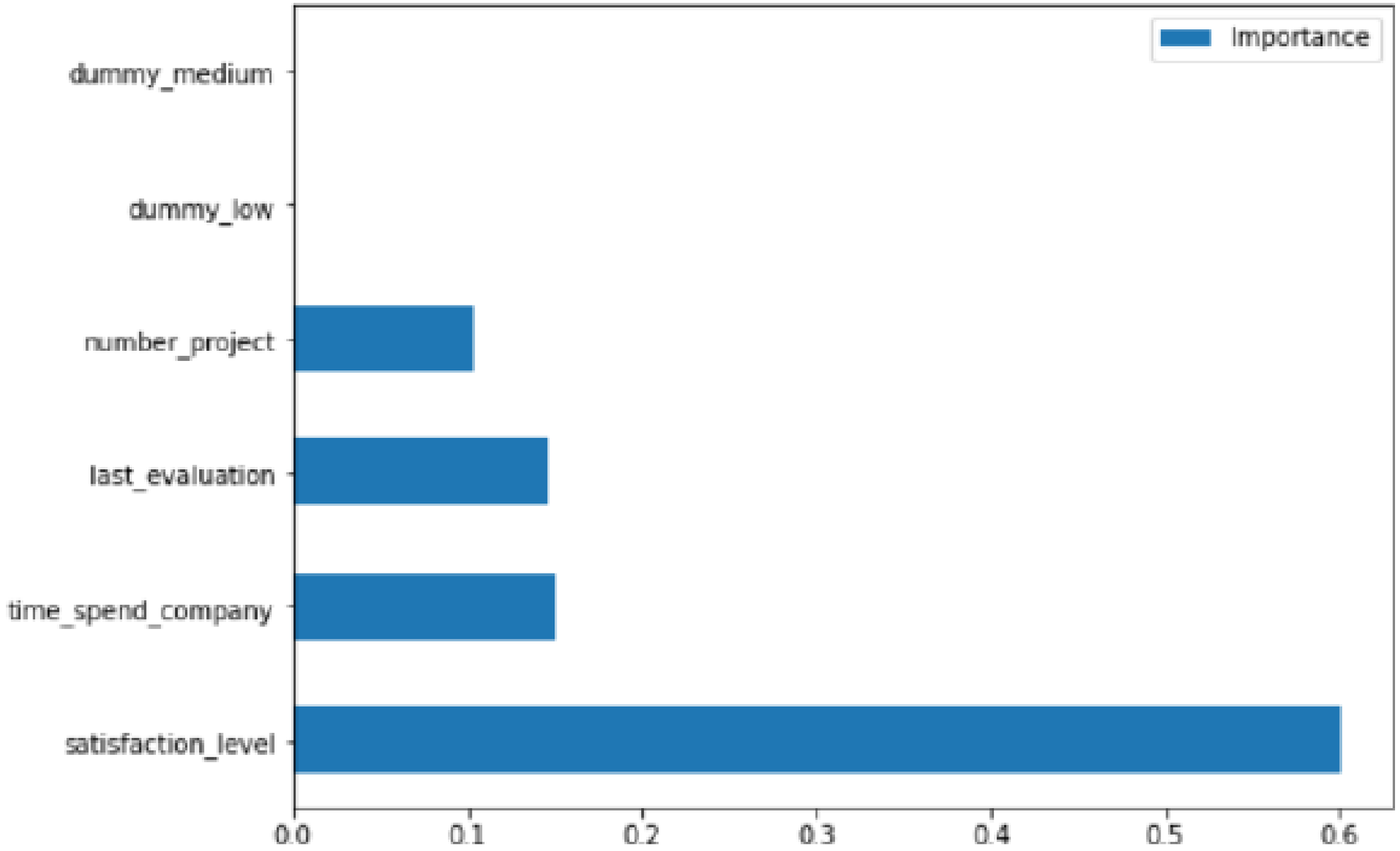
Methodology: Classification tree



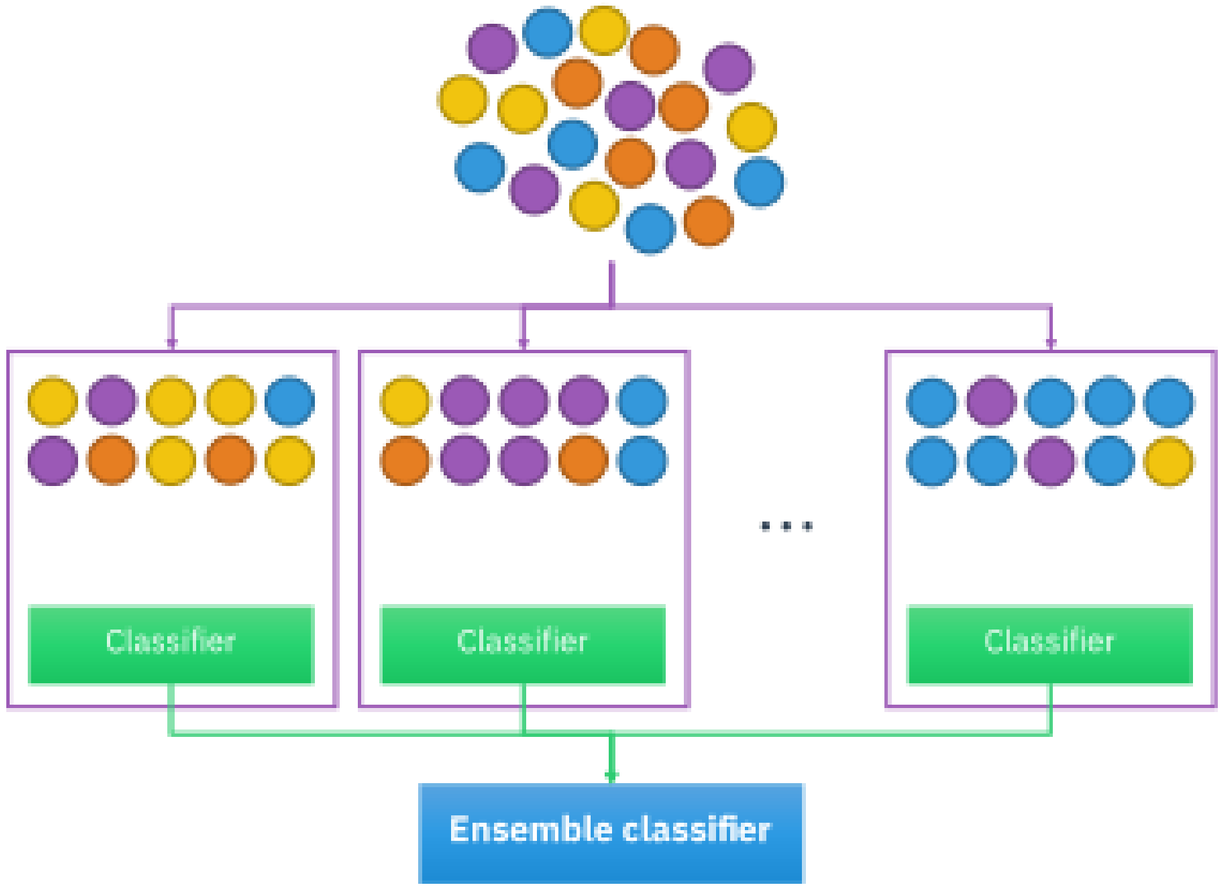
- **accuracy - 0.862**
- **specificity - 0.897**
- **sensitivity - 0.455**

Methodology: Classification tree

Feature importancies of independent variables



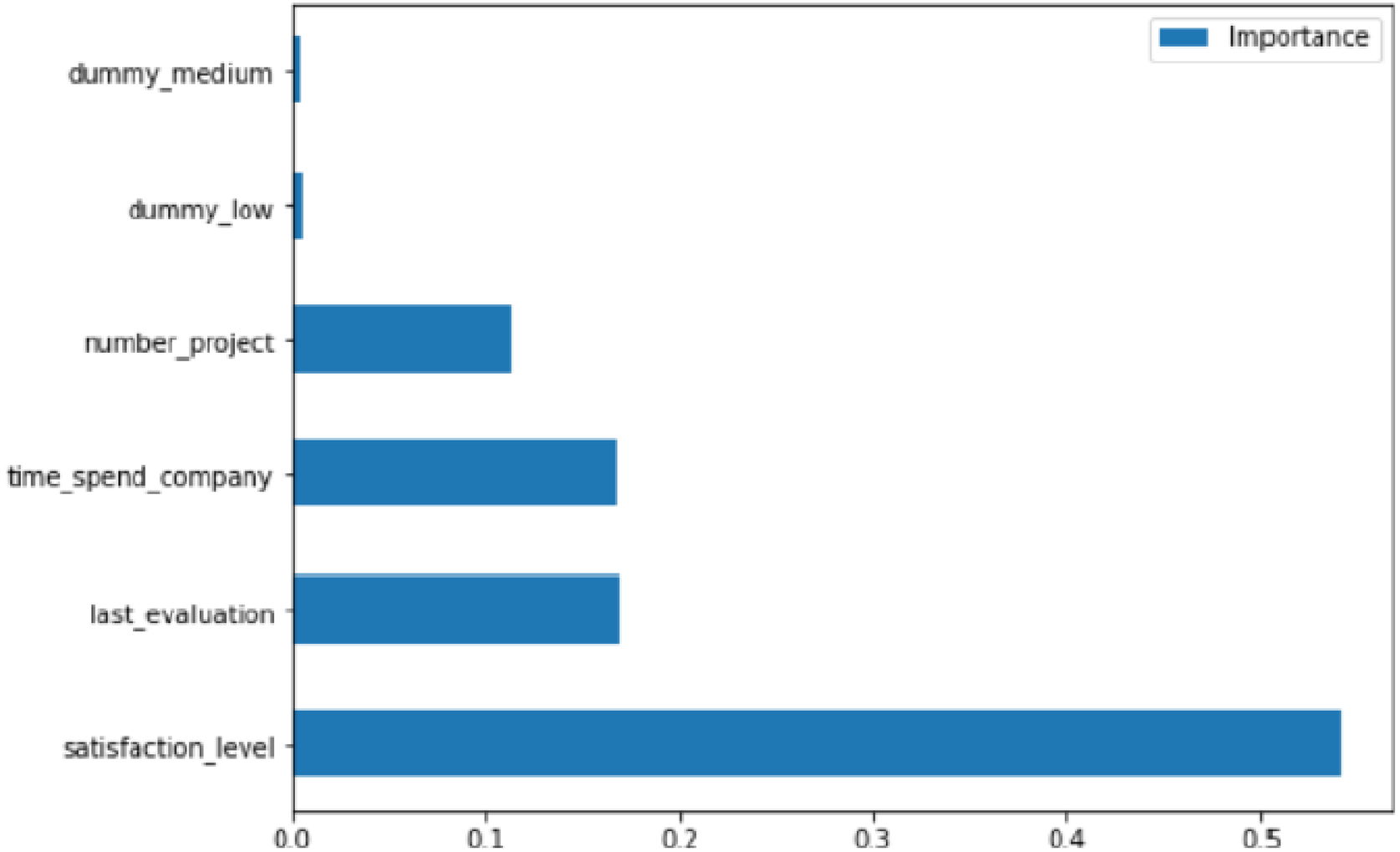
Methodology: Bagging



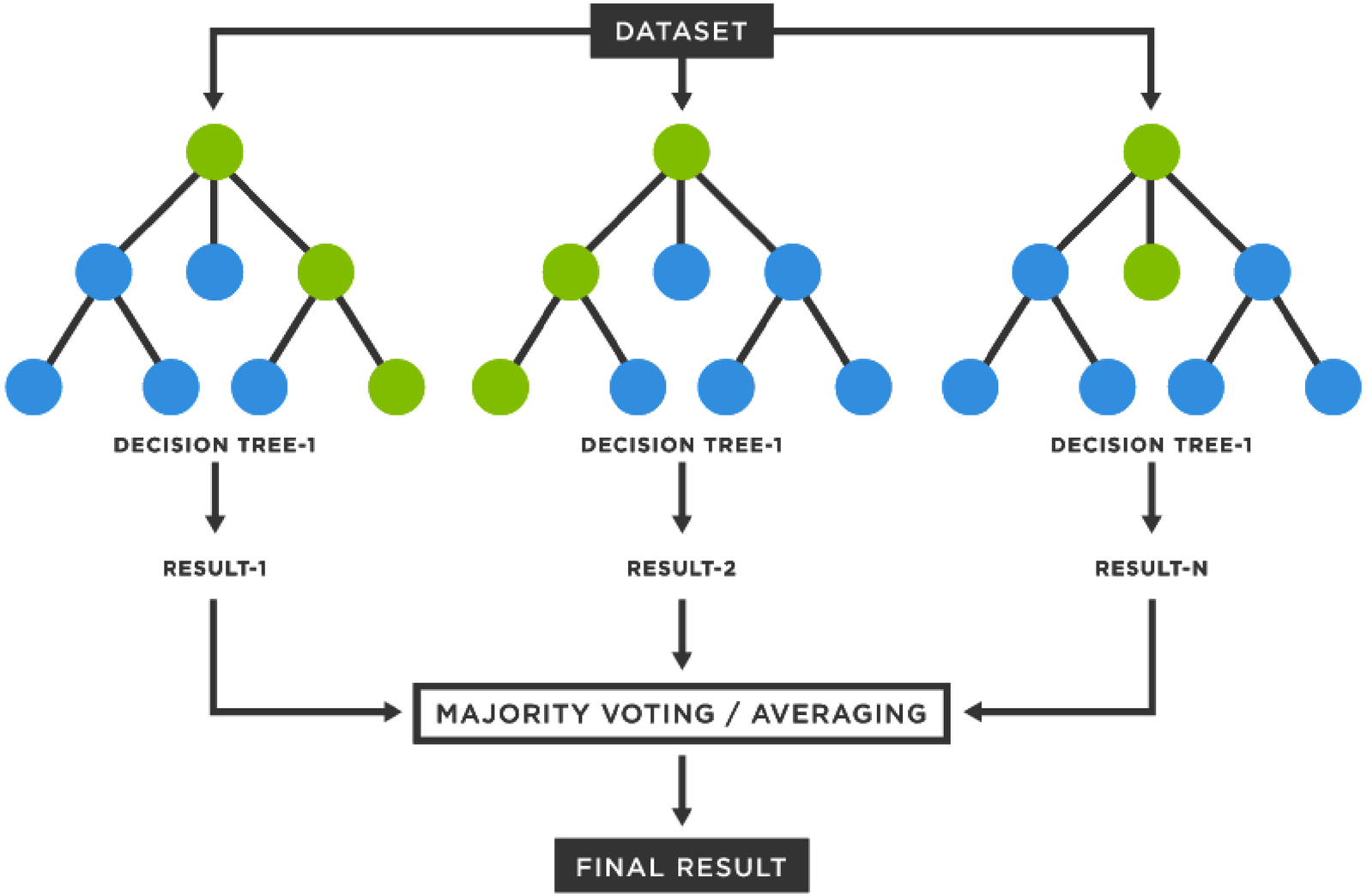
- **specificity - 0.974**
- **sensitivity - 0.971**

Methodology: Classification tree: Bagging

Feature importancies of independent variables



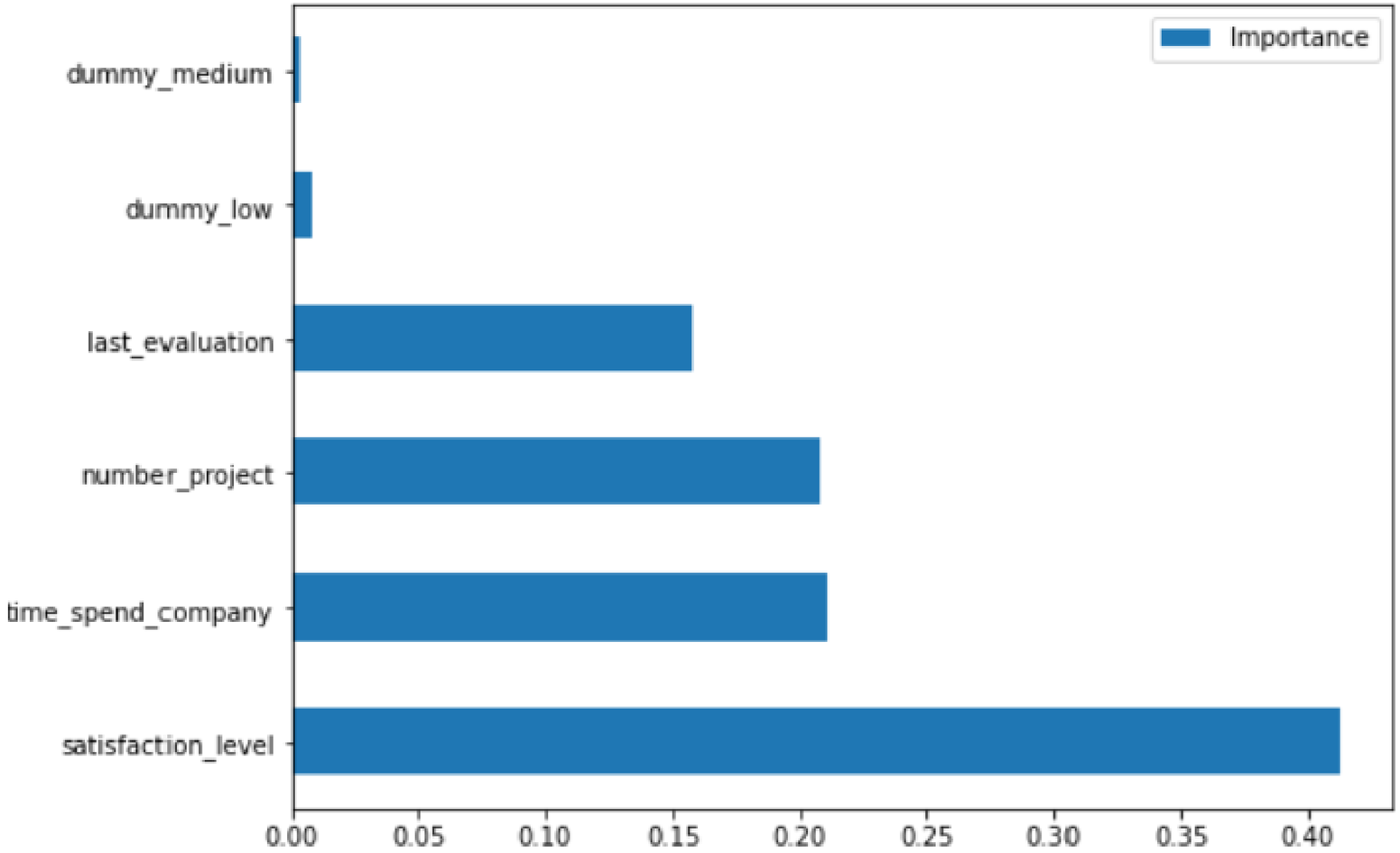
Methodology: Random Forest



- **specificity - 0.977**
- **sensitivity - 0.968**

Methodology: Classification tree: Random Forest

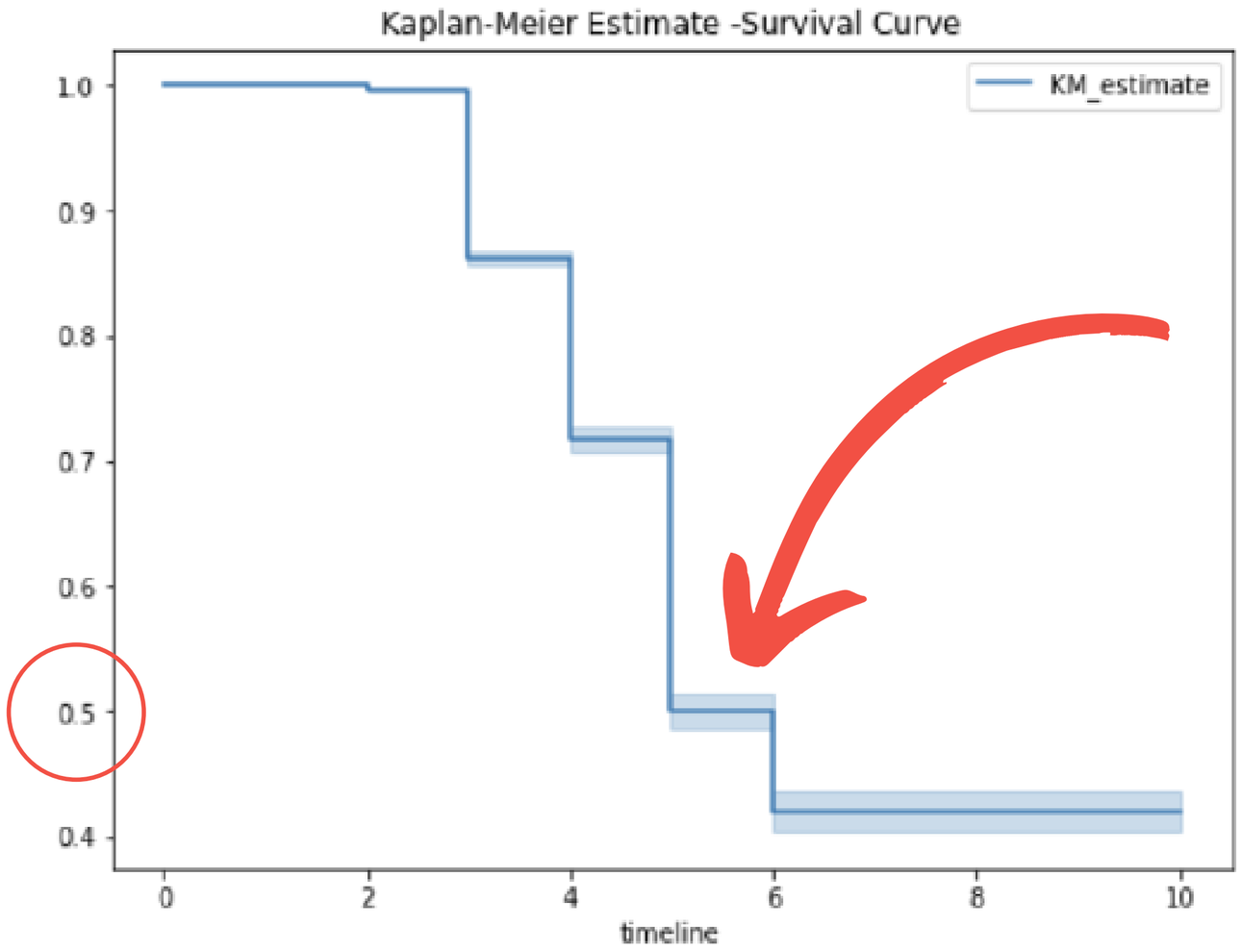
Feature importancies of independent variables



Methodology

Survival Analysis

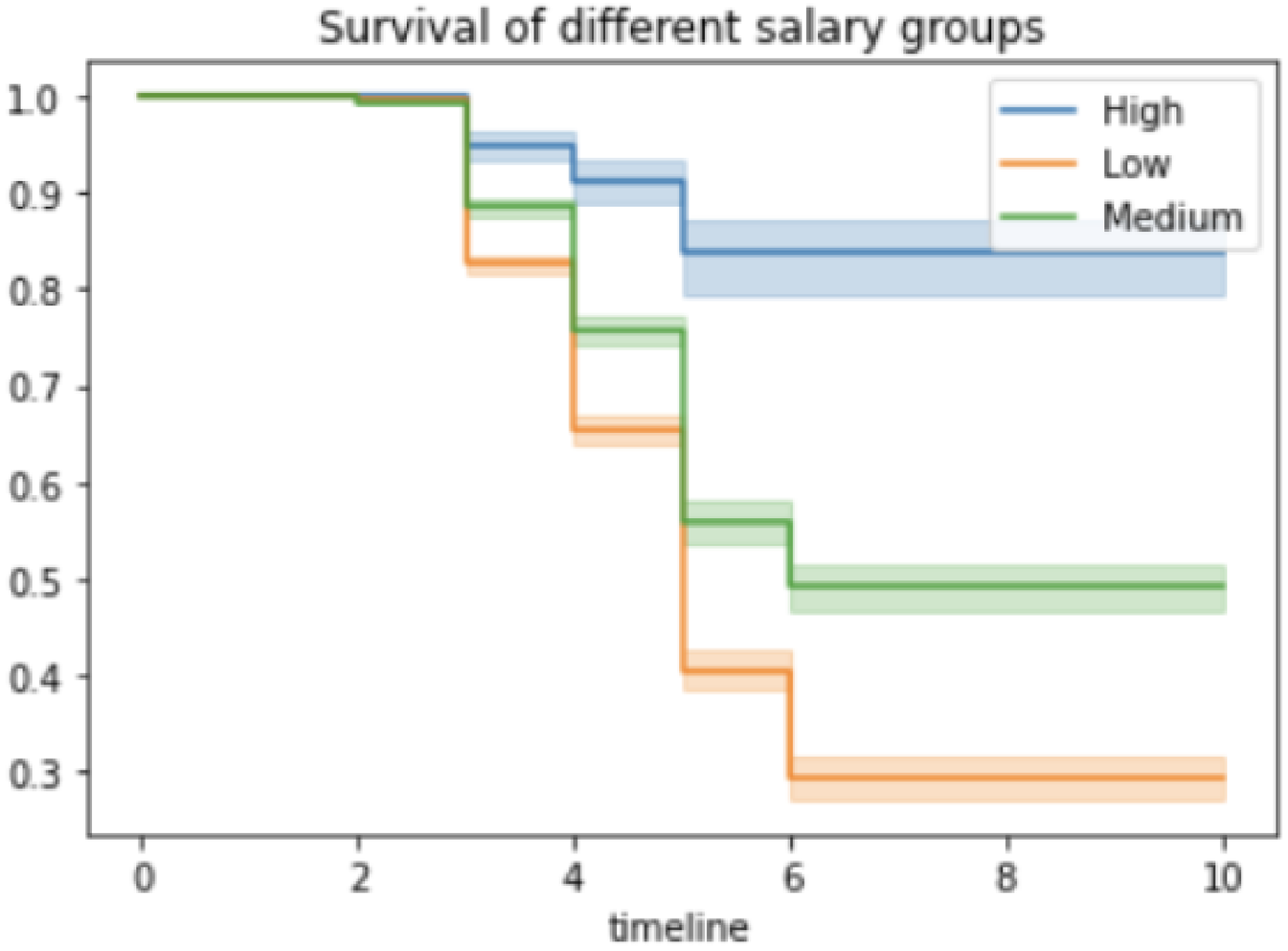
Methodology: Survival Analysis



Kaplan Meier Estimator

Median Survival Time
5-6 years

Methodology: Survival Analysis



Kaplan Meier Estimator

The probability of survival is the lowest for those who have low salary, and is the highest for those whose salary is high.

Methodology: Survival Analysis

Long Rank Test

Ho: there is no difference in survival curves

t_0	-1
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

	test_statistic	p	-log2(p)
0	177.06	<0.005	131.79

Low - medium

t_0	-1
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

	test_statistic	p	-log2(p)
0	291.83	<0.005	214.93

Low - high

t_0	-1
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

	test_statistic	p	-log2(p)
0	125.35	<0.005	94.24

Medium - High

Methodology: Survival Analysis

Long Rank Test

Ho: there is no difference in survival curves

Survival curves are statistically significantly different.

t_0	-1
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

	test_statistic	p	-log2(p)
0	177.06	<0.005	131.79

Low - medium

t_0	
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

	test_statistic	p	-log2(p)
0	291.83	<0.005	214.93

Low - high

t_0	-1
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

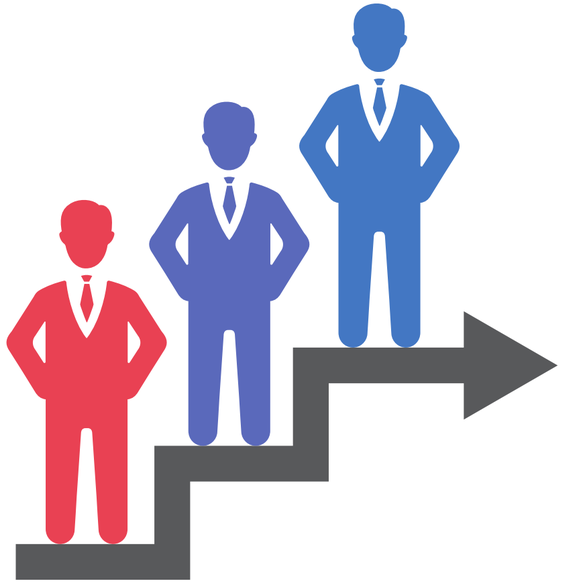
	test_statistic	p	-log2(p)
0	125.35	<0.005	94.24

Medium - High

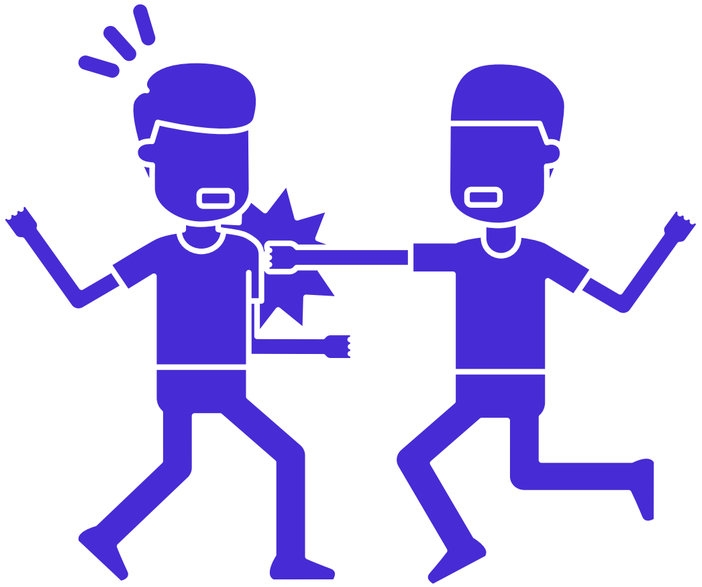
Conclusion and Suggestions

Pay more!

Appreciate!



Create challenges!



THANK YOU