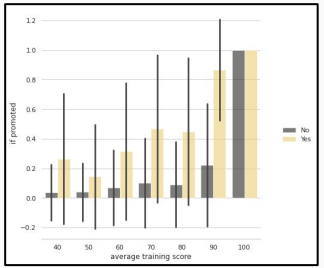
HR PREDICTION ANALYSIS FOR PROMOTIONS

Group 9



	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	award
0	8	7	3	2	2	1	35	5.0	8	
1	5	22	1	1	1	1	30	5.0	4	
2	8	19	1	1	2	1	34	3.0	7	
3	8	23	1	1	1	2	39	1.0	10	
4	9	26	1	1	1	1	45	3.0	2	
...	...	...	...	...	...	...	...	...	...	
54802	8	14	1	1	1	2	31	1.0	2	
54803	9	14	1	1	2	1	48	3.0	17	
54804	5	27	3	2	1	1	37	2.0	6	
54805	1	1	1	1	1	1	27	5.0	3	
54807	3	22	1	1	1	1	27	1.0	5	

48660 rows × 12 columns

THE BOTTOM LINE

PROMOTION CYCLE

Pro

SOLUTION:
PREDICTION ANALYSIS

AGENDA

01 DATA

02 FEATURE SELECTION

03 MODELS

04 APPLICATION

01.
DATA

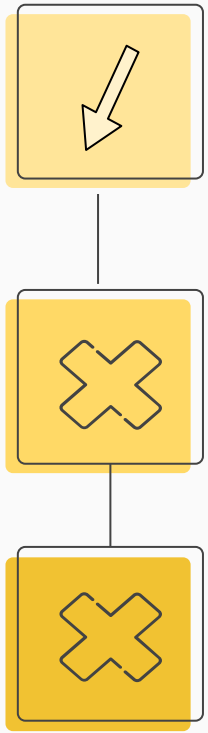
	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service
0	65438	Sales & Marketing	region_7	Master's & above	f	sourcing	1	35	5.0	8
1	65141	Operations	region_22	Bachelor's	m	other	1	30	5.0	4
2	7513	Sales & Marketing	region_19	Bachelor's	m	sourcing	1	34	3.0	7
3	2542	Sales & Marketing	region_23	Bachelor's	m	other	2	39	1.0	10
4	48945	Technology	region_26	Bachelor's	m	other	1	45	3.0	2
...	...	...	...	...	...	...	...	...	...	...
54803	3030	Technology	region_14	Bachelor's	m	sourcing	1	48	3.0	17
54804	74592	Operations	region_27	Master's & above	f	other	1	37	2.0	6
54805	13918	Analytics	region_1	Bachelor's	m	other	1	27	5.0	3
54806	13614	Sales & Marketing	region_9	NaN	m	sourcing	1	29	1.0	2
54807	51526	HR	region_22	Bachelor's	m	other	1	27	1.0	5

54808 rows × 13 columns

	department	regi	ucation	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	award
0	8		3	2	2	1	35	5.0	8	
1	5		1	1	1	1	30	5.0	4	
2	8		1	1	2	1	34	3.0	7	
3	8		1	1	1	2	39	1.0	10	
4	9	6	1	1	1	1	45	3.0	2	
...			...	...	...	...	...	...	...	
54802			1	1	1	2	31	1.0	2	
54803		14	1	1	2	1	48	3.0	17	
54804	5	27	3	2	1	1	37	2.0	6	
54805	1	1	1	1	1	1	27	5.0	3	
54807	3	22	1	1	1	1	27	1.0	5	

48660 rows × 12 columns

CLEANING DATA



STEP 01

Categorical to Numerical

STEP 02

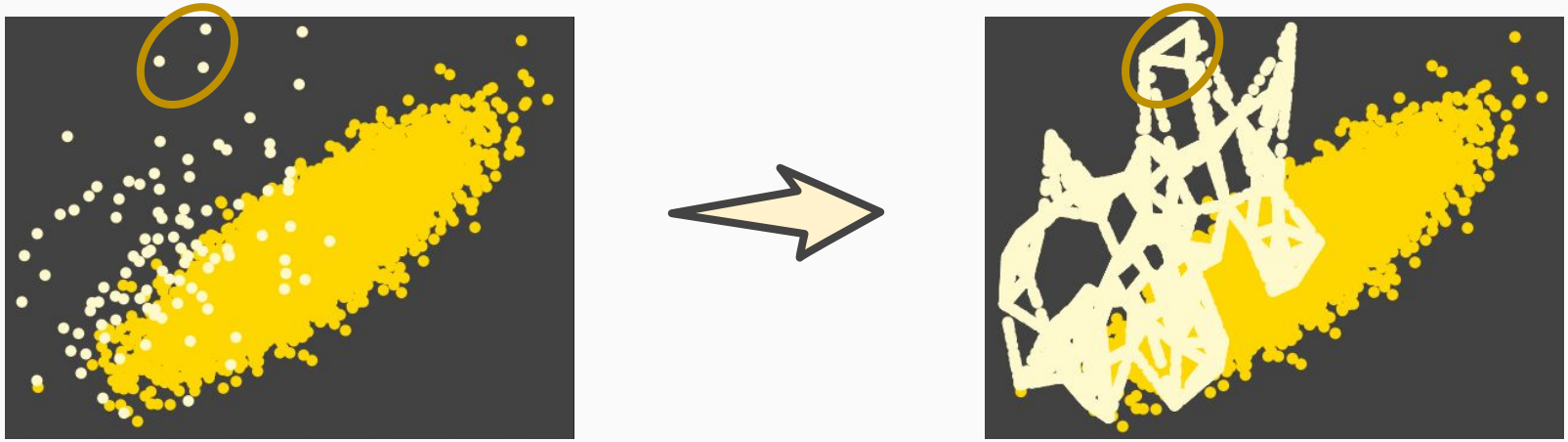
Removed Employee ID's

STEP 03

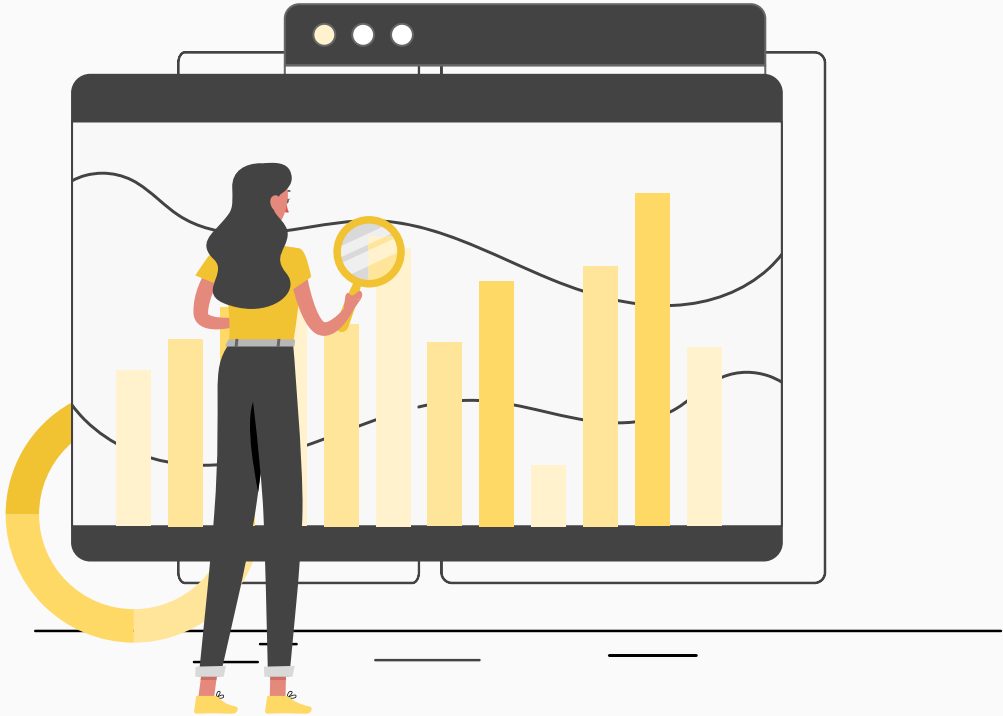
Removed NaN's and blanks

HANDLING IMBALANCED DATA - SMOTE

NOT PROMOTED	PROMOTED	
44428	4232	BEFORE SMOTE
44428	44428	AFTER SMOTE

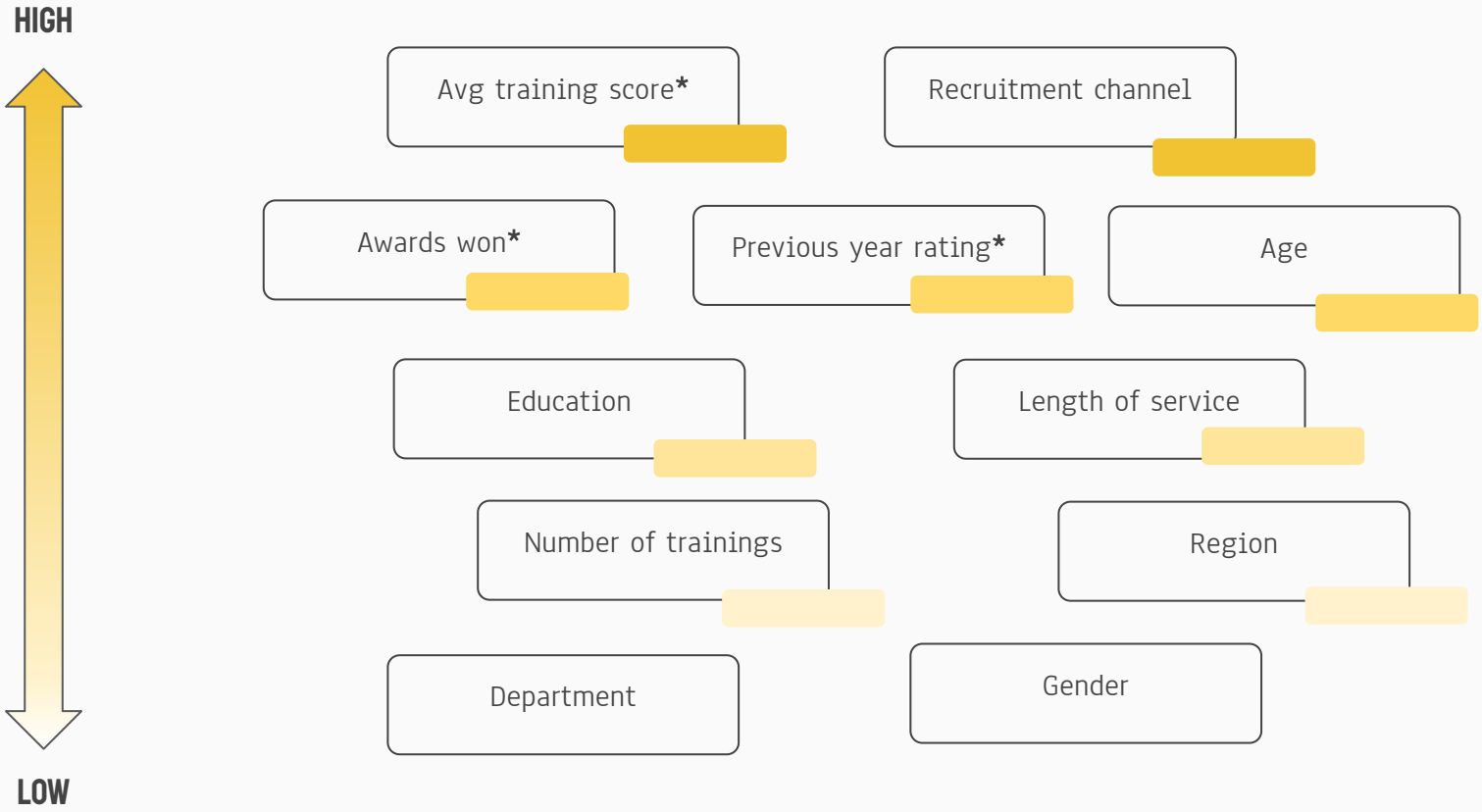


02.
FEATURE SELECTION

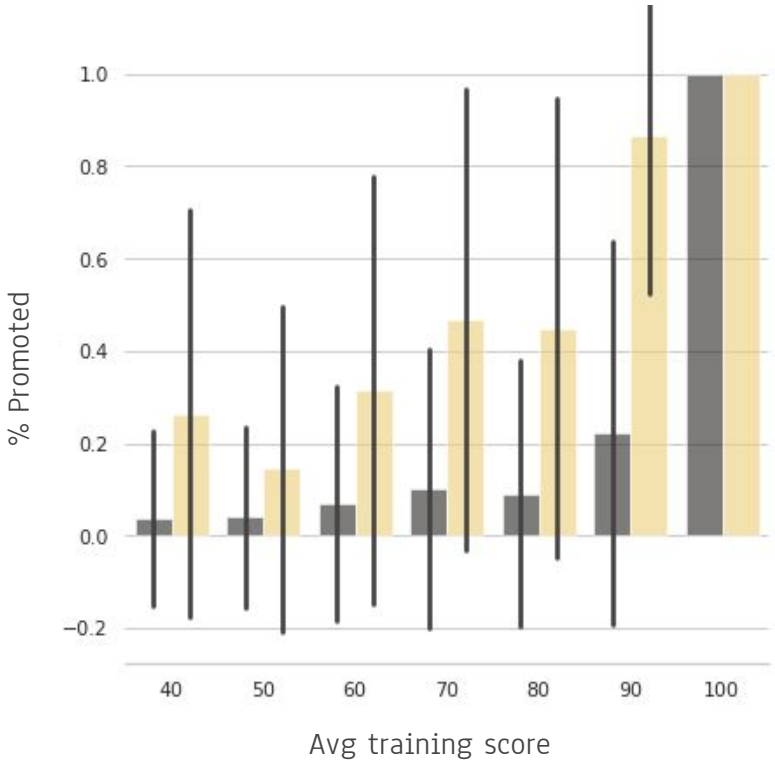


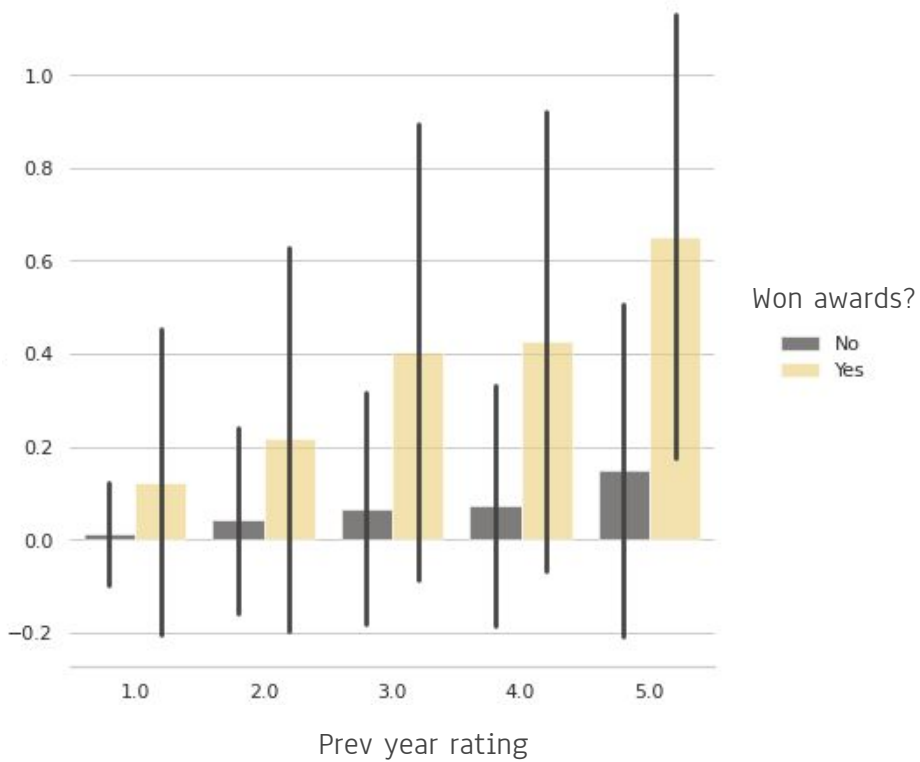
(*) indicates top 3 chi-squared statistics

FEATURE INFLUENCE

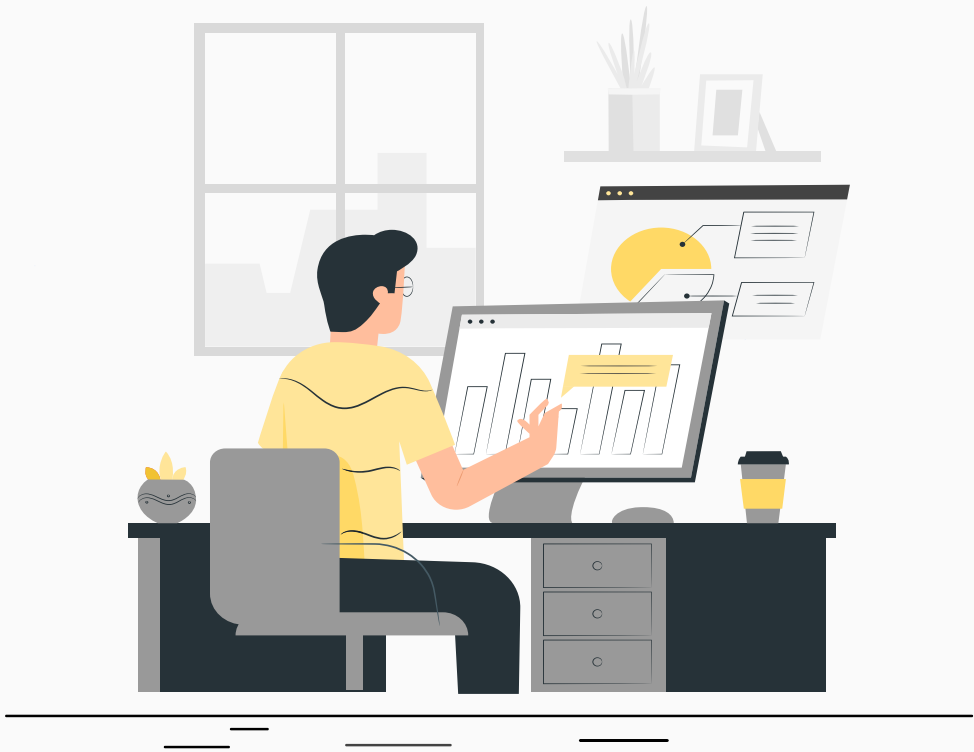


FEATURE VISUALIZATION





03.
MODELS



MODEL COMPARISON

	INTERPRET-ABILITY	COMPUTE SPEED	RECALL	PRECISION
LOGISTIC REGRESSION	High	Fast	68.5%	70.4%
LINEAR SVM	Avg	Fast	68.3%	72.4%
POLYNOMIAL SVM	Low	Slow	72.8%	75.7%
RBF SVM	Low	Slow	75.8%	80.3%
NAIVE BAYES	High	Fast	65.8%	77.4%
MULTI-LAYER PERCEPTRON	Low	Slow	78.9%	60.7%

- Precision mitigates false positives (desired)
- Features removed: Department and Gender
- Used Cross Validation

04.
APPLICATION

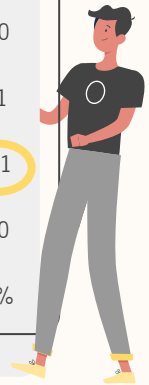
(*) indicates top 3 chi-squared statistics

PERSON A

Gender	Female
Age	36
Department	R&D
Education	MS & Above
Recruitment	Sourcing
Length of service	9
Awards won *	0
Number of trainings	1
Avg training score *	80
Prev year rating *	5.0/5.0
Prediction (Yes)	76.8%

PERSON B

Gender	Male
Age	60
Department	Technology
Education	MS & Above
Recruitment	Sourcing
Length of service	9
Awards won *	0
Number of trainings	1
Avg training score *	61
Prev year rating *	5.0/5.0
Prediction (Yes)	19.4%

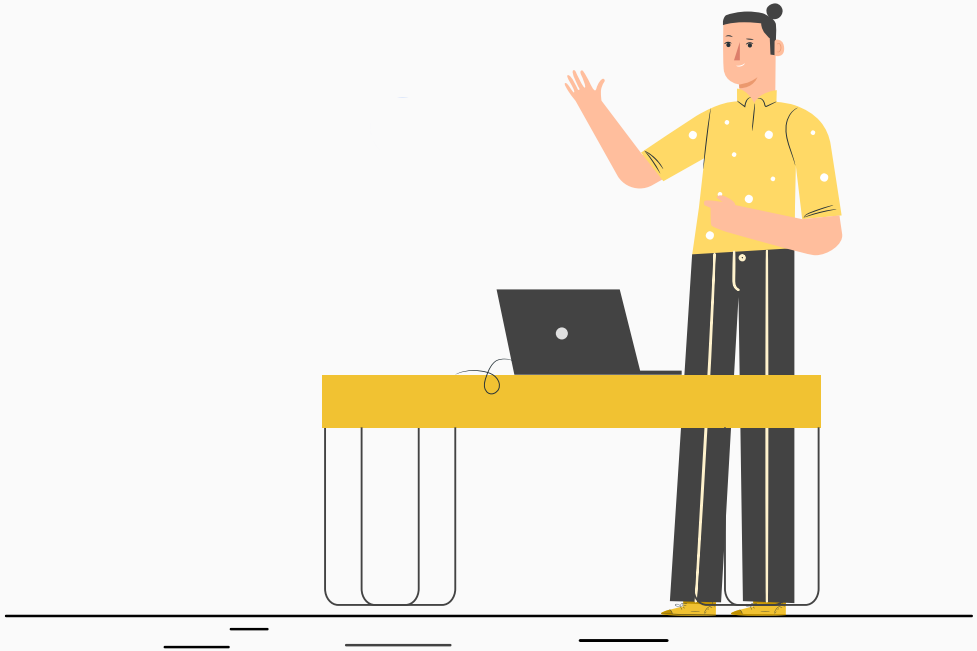


PERSON C

Gender	Male
Age	34
Department	Analytics
Education	BS
Recruitment	Sourcing
Length of service	7
Awards won *	1
Number of trainings	1
Avg training score *	90
Prev year rating *	5.0/5.0
Prediction (Yes)	100%

LIMITATIONS AND NEXT STEPS

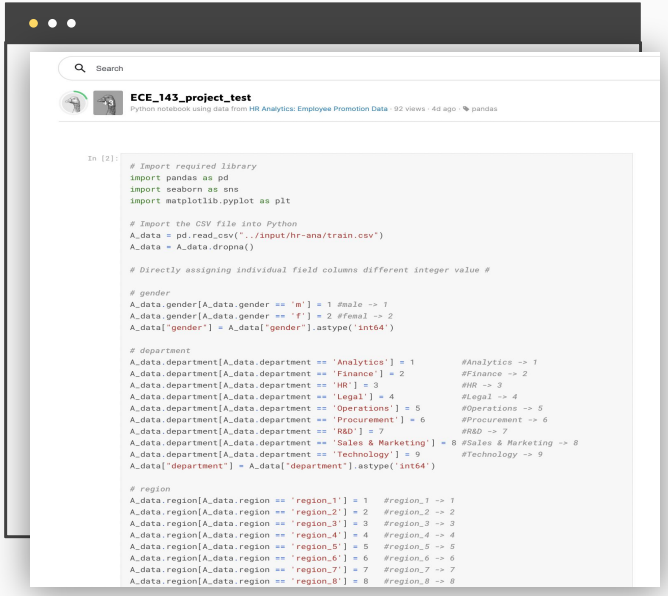
- Lack of identifiable information
- Need more data/datasets
- Productize codebase



THANKS

Does anyone have any questions?

Group 9



Kaggle Notebook: https://www.kaggle.com/eugeneli9/ece-143-project-test