IT Salaries 2.0

Lelia Erscoi

Agenda

About the Data

KNN - About

KNN - Results

LDA - About

LDA - Results

Logistic Regression - About

Logistic Regression - Results

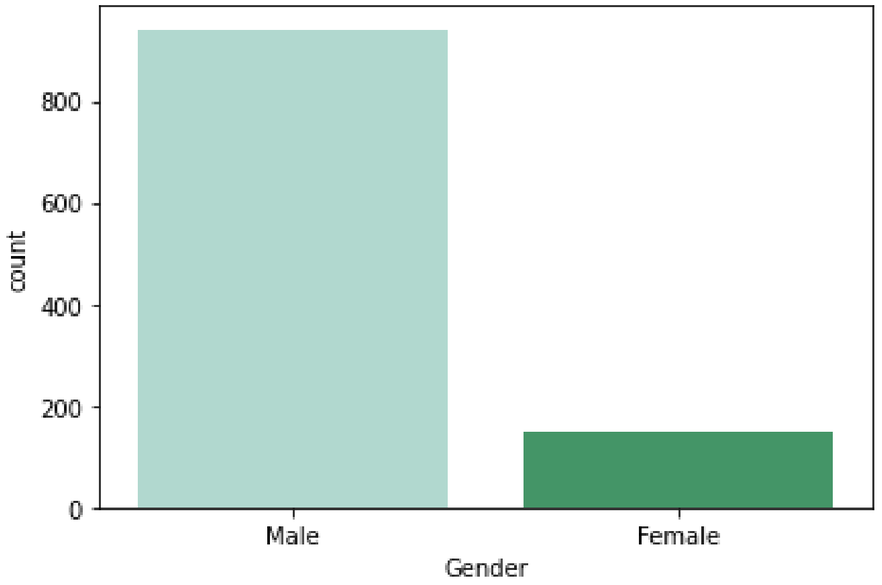
Model Comparison

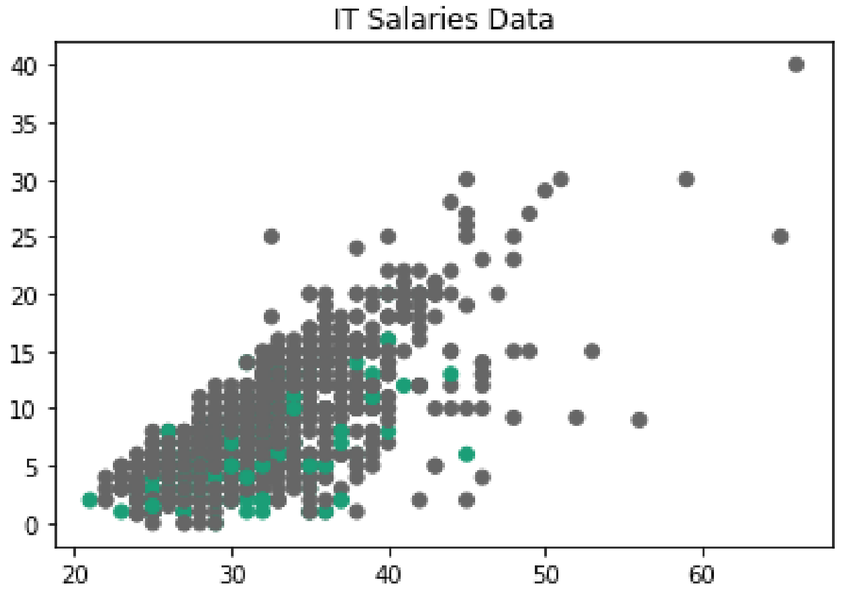
Feature Relevance

(More) About the Data

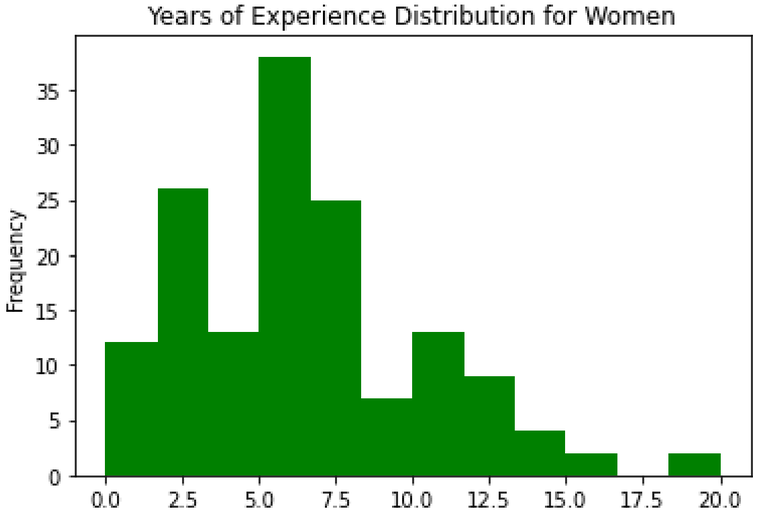
Insights:

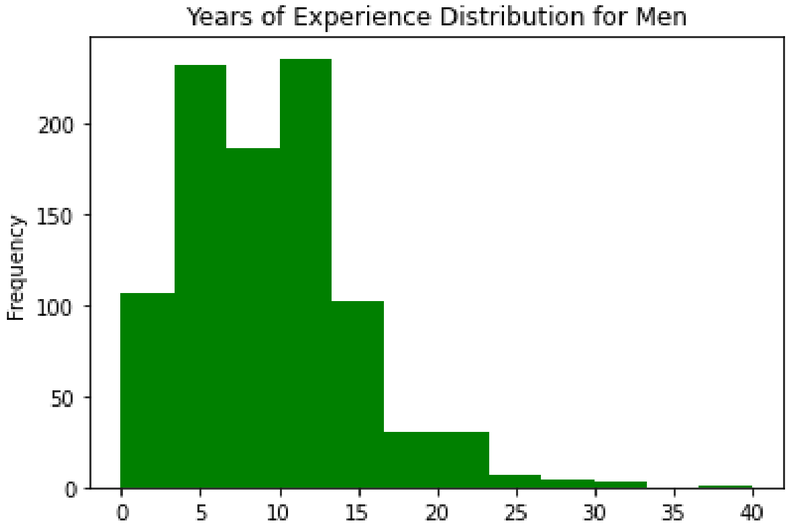
- 1092 x 20 data
- Focus on the Gender dimension
- 151 "female" and 940 "male"
- Third gender self identity option, "Diverse", severely underrepresented

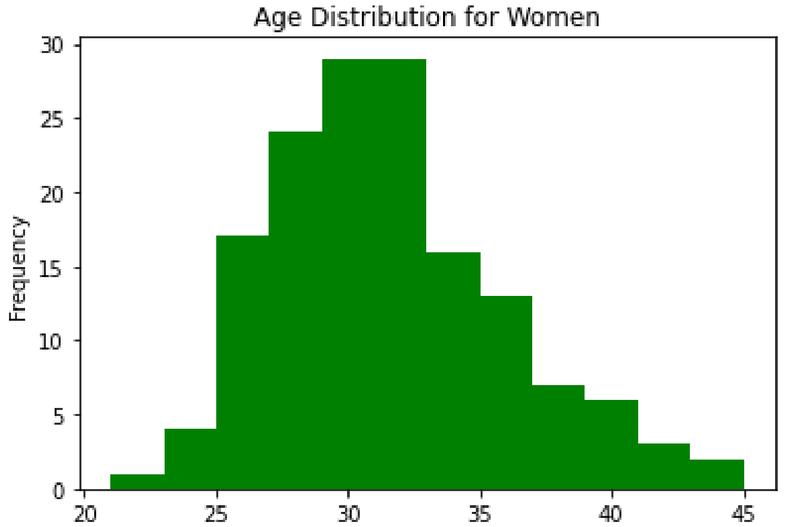


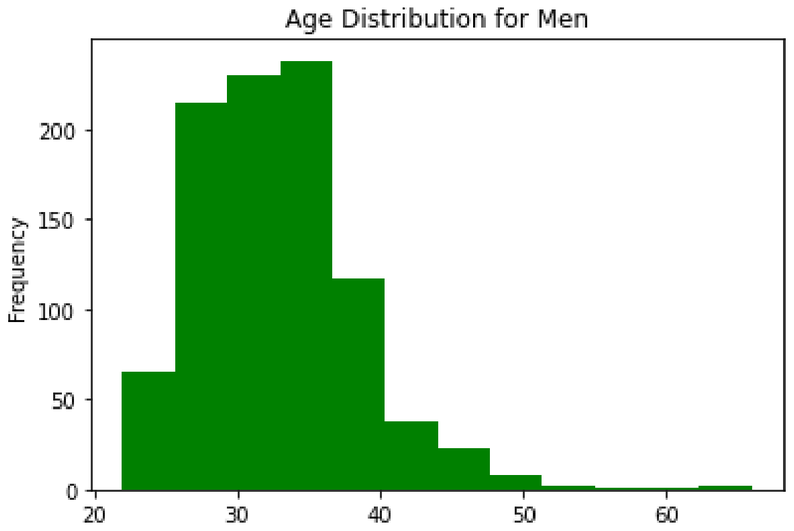


Distributions By Gender







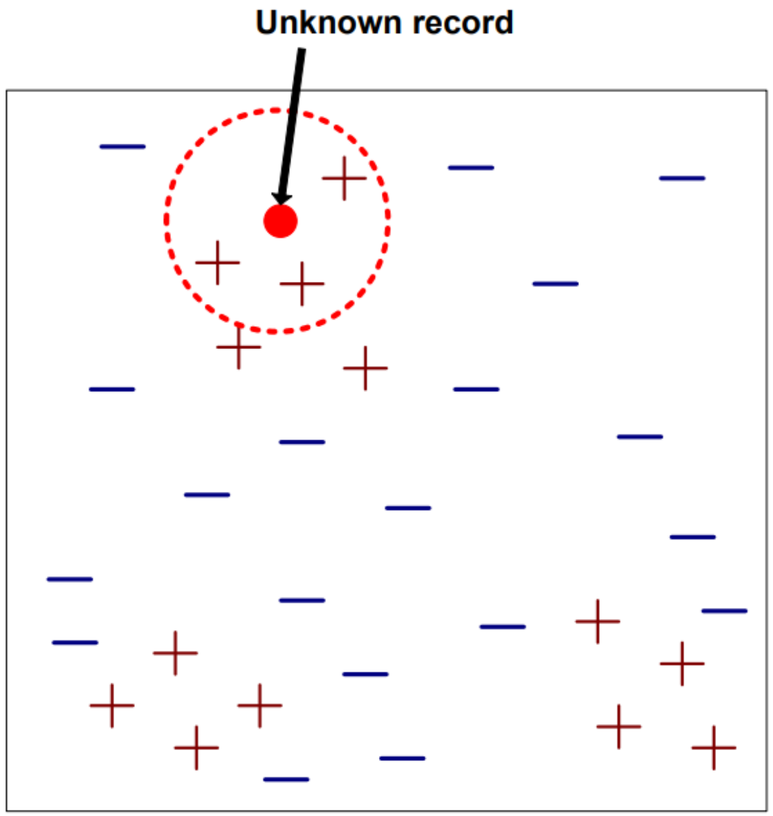


Insights:

- Years of experience
  - Mean: 6.37 F, 9.14 M
  - Max: 20 F, 40 M
- Age:
  - Mean: 31.12 F, 32.64 M
  - Min: 21 F, 22 M
  - Max: 45 F, 66 M

K Nearest Neighbors

About



Expectations

Lazy Learner

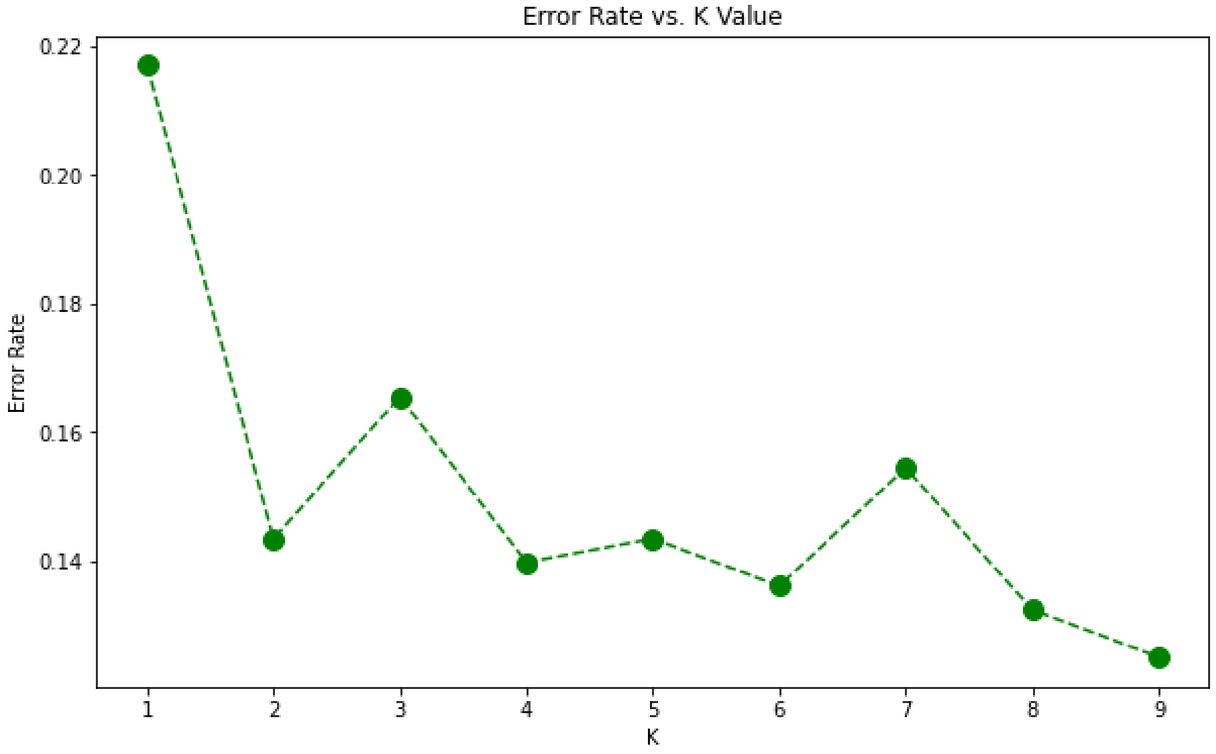
Sensitive to noise

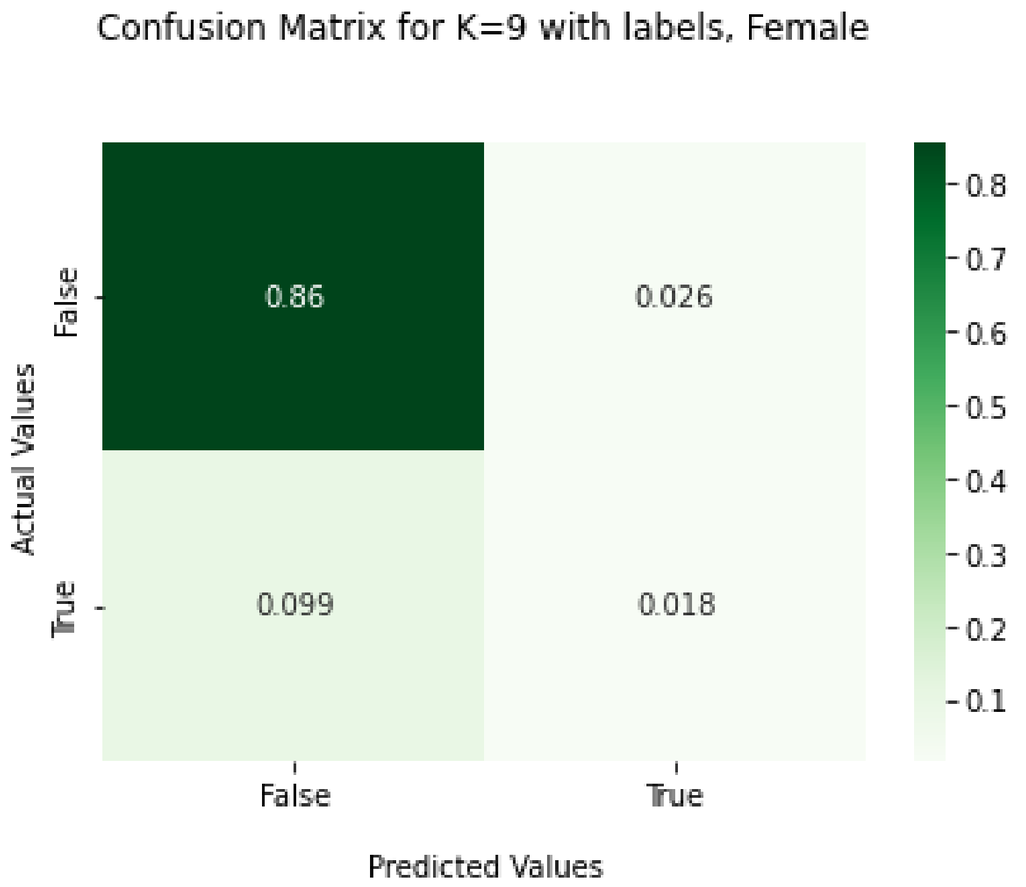
Works better when attributes are scaled

K Nearest Neighbors

Results (Female Label)

Metric	Accuracy	Precision	Recall	F1 Score
Score	87.5%	41.66%	15.62%	22.72%

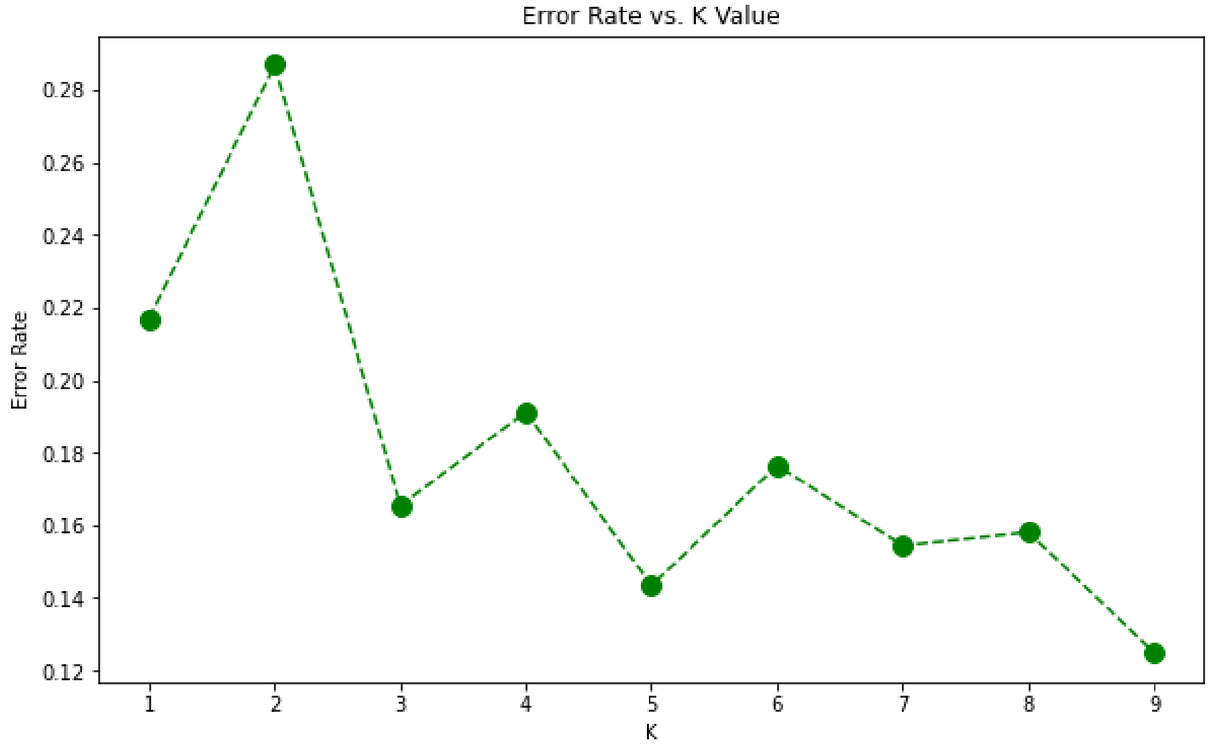


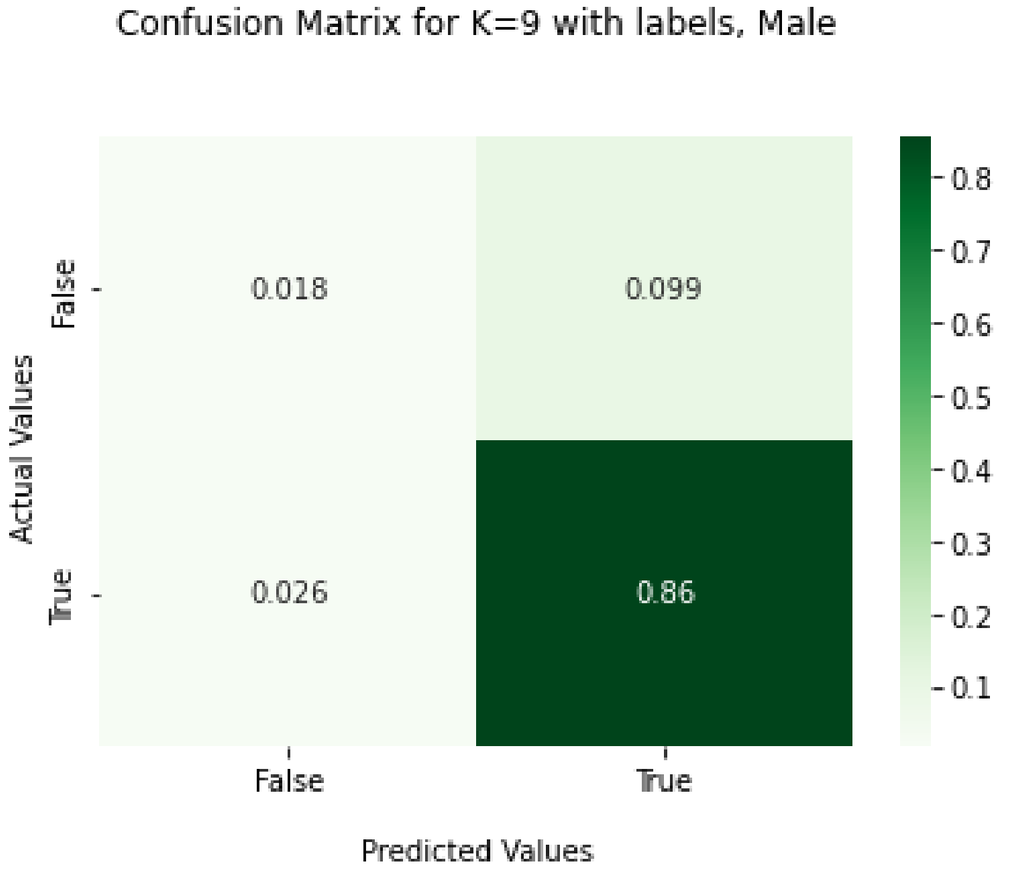


K Nearest Neighbors

Results (Male Label)

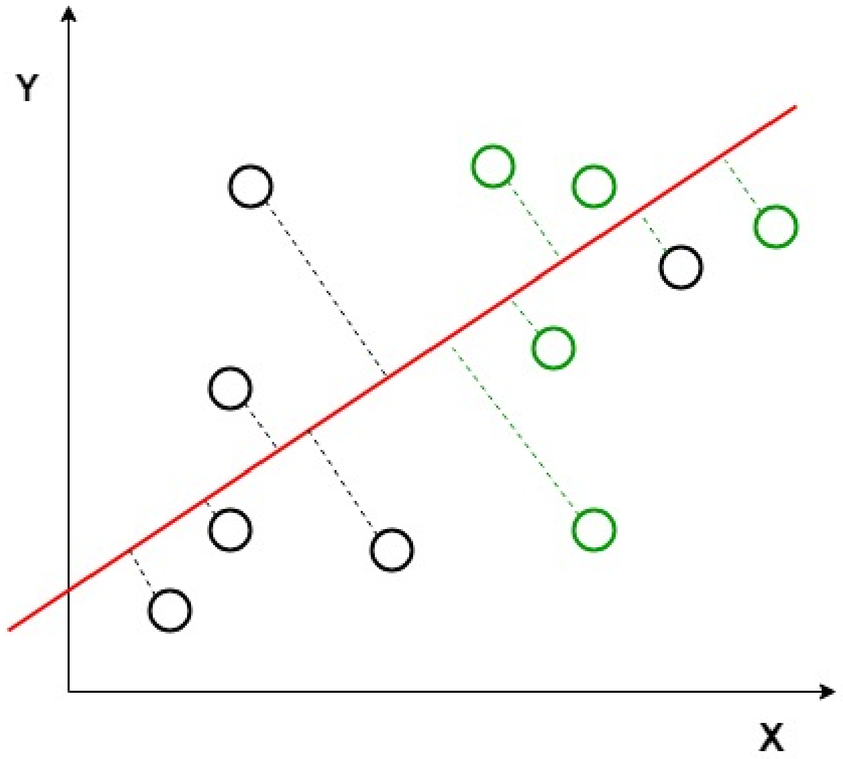
Metric	Accuracy	Precision	Recall	F1 Score
Score	84.24%	89.61%	97.08%	93.19%





Linear Discriminant Analysis

About



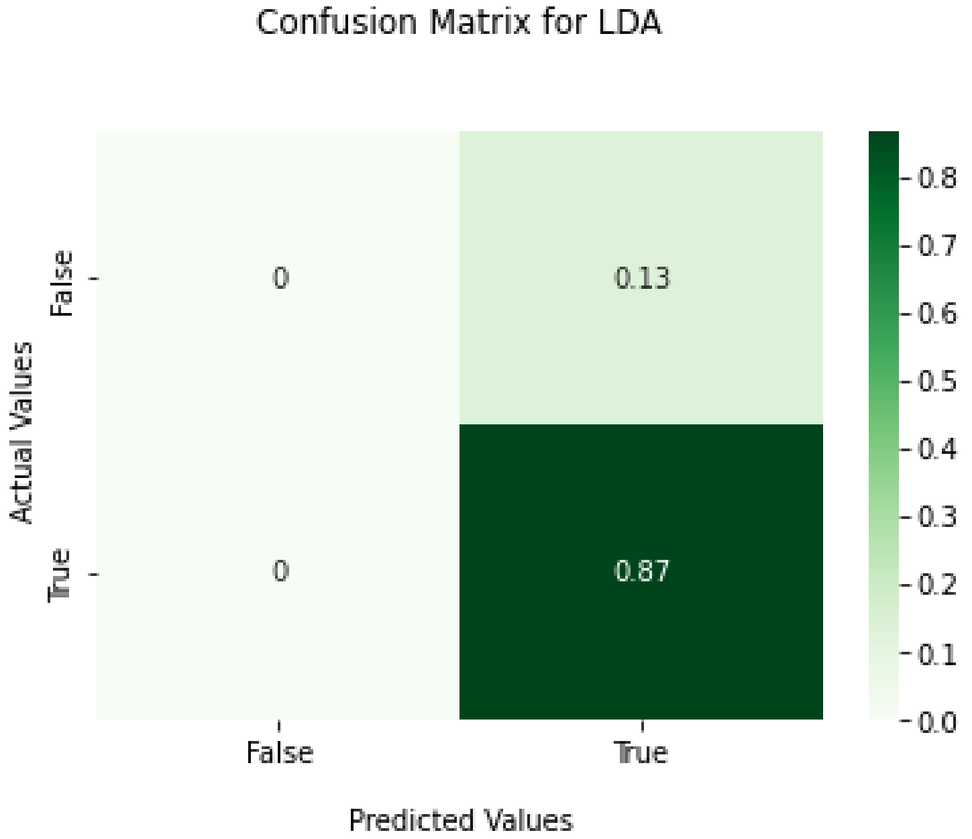
Expectations

Dimensionality reduction technique
Works if data is normally distributed and
If the covariance matrices of the classes are equal

LDA

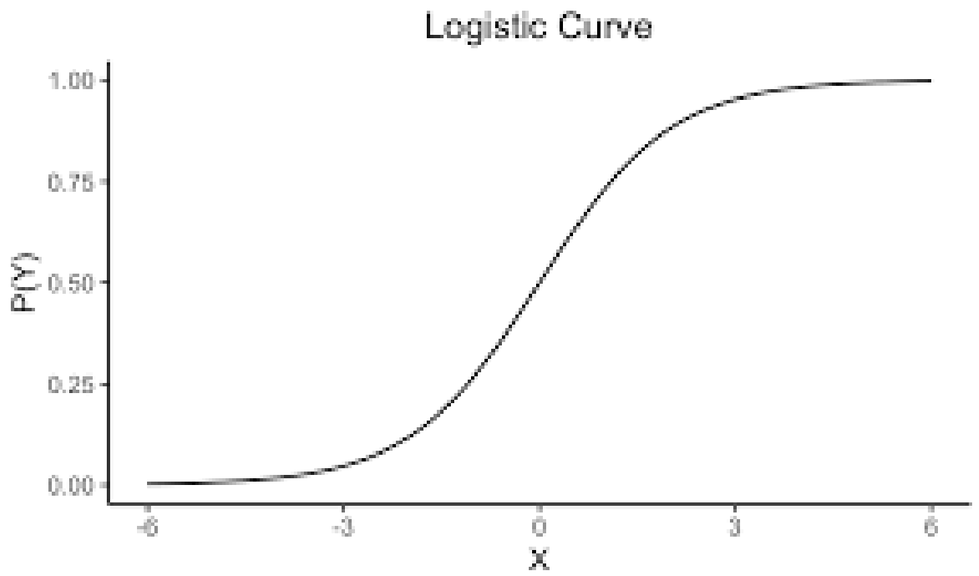
Results (Male Label)

Metric	Accuracy	Precision	Recall	F1 Score
Score	86.69%	86.69%	100%	92.87%



Logistic Regression

About



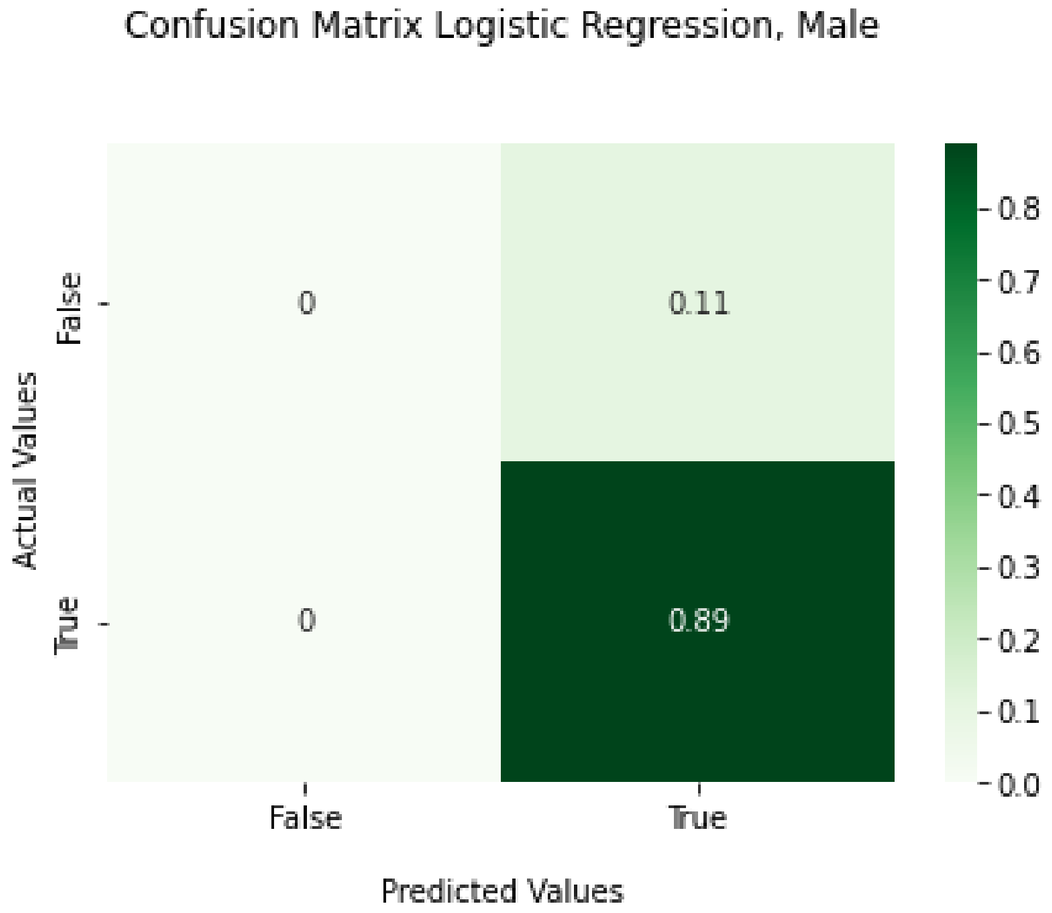
Expectations

Non-linear decision boundaries
Loss function more difficult to compute

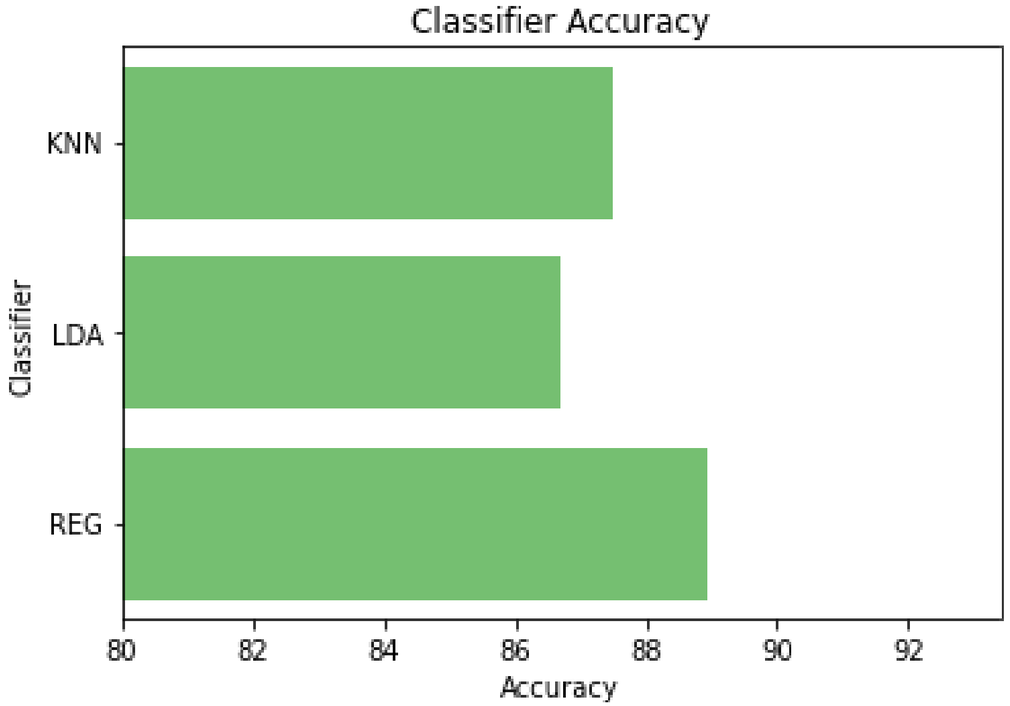
Logistic Regression

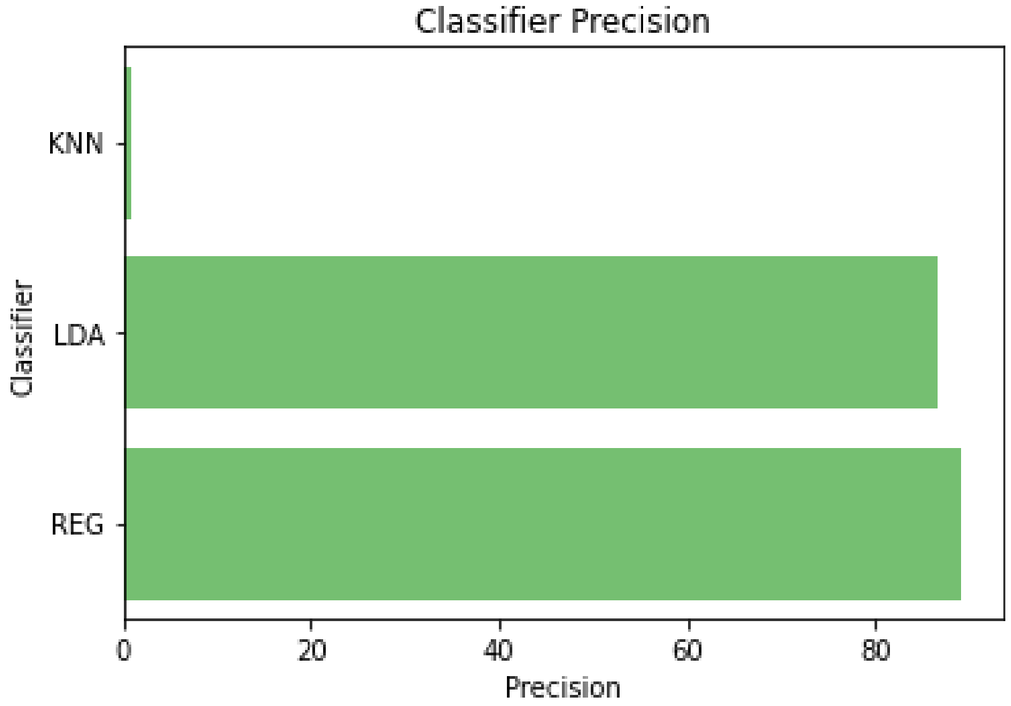
Results (Male Label)

Metric	Accuracy	Precision	Recall	F1 Score
Score	86.69%	88.95%	100%	94.15%

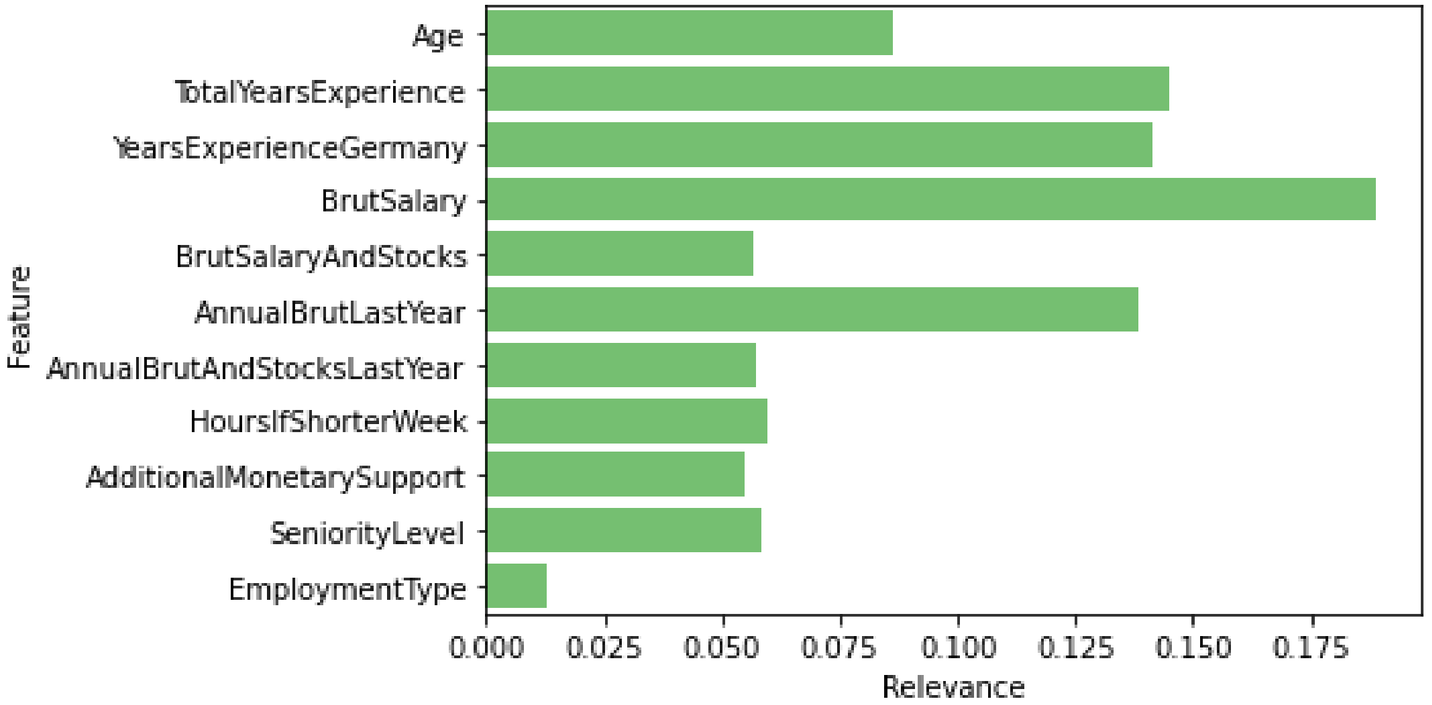


Model Comparison



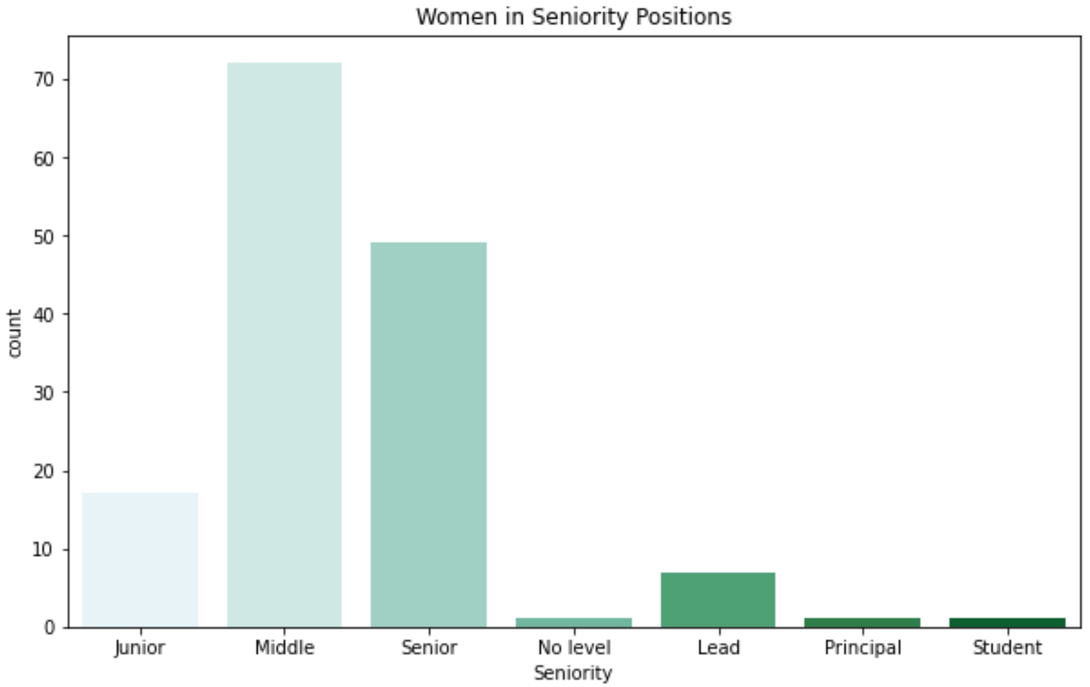


Feature Relevance



Findings


Men in Seniority Positions
count
0
100
200
300
400
Senior
Junior
Middle
Lead
Head
VP
Manager
Director
Principal
Intern
Seniority


Women in Seniority Positions
count
0
10
20
30
40
50
60
70
Junior
Middle
Senior
No level
Lead
Principal
Student
Seniority

Findings

Logit Regression Results

Dep. Variable:	y	No. Observations:	760
Model:	Logit	Df Residuals:	749
Method:	MLE	Df Model:	10
Date:	Sat, 15 Oct 2022	Pseudo R-squ.:	0.1011
Time:	20:01:25	Log-Likelihood:	-287.21
converged:	True	LL-Null:	-319.52
Covariance Type:	nonrobust	LLR p-value:	4.816e-10

	coef	std err	z	P>|z|	[0.025	0.975]
Age	-0.0198	0.023	-0.860	0.390	-0.065	0.025
TotalYearsExperience	0.1344	0.035	3.812	0.000	0.065	0.204
YearsExperienceGermany	-0.0455	0.040	-1.151	0.250	-0.123	0.032
BrutSalary	1.489e-05	9.53e-06	1.563	0.118	-3.78e-06	3.36e-05
BrutSalaryAndStocks	1.851e-05	9.94e-06	1.862	0.063	-9.74e-07	3.8e-05
AnnualBrutLastYear	1.175e-05	1.06e-05	1.108	0.268	-9.03e-06	3.25e-05
AnnualBrutAndStocksLastYear	6.952e-06	9.15e-06	0.760	0.447	-1.1e-05	2.49e-05
HoursIfShorterWeek	0.0017	0.013	0.129	0.898	-0.024	0.027
AdditionalMonetarySupport	-1.445e-05	0.001	-0.023	0.981	-0.001	0.001
SeniorityLevel	-0.1080	0.183	-0.592	0.554	-0.466	0.250
EmploymentType	-0.0430	0.161	-0.267	0.790	-0.359	0.273

Thank You!