Main Page for Pipey

Select data file and init proc

preprocessor

Select Encoding Strategy

Imputation

Use Selected Encoding Steps to Encode the Train and Test data

Run and Score Models

EDA and graphs

Optimize Run and Score

pygwalker

👋 Run and Score Page with no optimization (Grid Search) 👋

Scores Accuracy, Balanced Accuracy, Precision and other Metrics

y_train X_train_scaled

	Attrition
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	1
8	0
9	0

	JobRole_Healthcare Representative	JobRole_Human Resources	JobRole_Labora
0	-0.3124	-0.1864	
1	-0.3124	-0.1864	
2	-0.3124	-0.1864	
3	-0.3124	-0.1864	
4	-0.3124	-0.1864	
5	3.2008	-0.1864	
6	-0.3124	-0.1864	
7	-0.3124	-0.1864	
8	-0.3124	-0.1864	
9	-0.3124	-0.1864	

y_test

X_test_scaled

	Attrition
0	0
1	0
2	1
3	0
4	0
5	0
6	0
7	0
8	0
9	0

	JobRole_Healthcare Representative	JobRole_Human Resources	JobRole_Labora
0	-0.3124	-0.1864	
1	-0.3124	-0.1864	
2	-0.3124	5.3651	
3	-0.3124	-0.1864	
4	-0.3124	-0.1864	
5	-0.3124	-0.1864	
6	-0.3124	-0.1864	
7	-0.3124	-0.1864	
8	-0.3124	-0.1864	
9	-0.3124	-0.1864	

Ready to go. Proceeding...

None

Voting Classifier Scoring

Type of Data Set	Accuracy	Balanced Accuracy
Training	91.38	75.08
Test	88.86	59.06

Individual Classifier Scoring (Accuracy and Balanced Accuracy)

Classifier Name	Training Accuracy	Training Balanced Accuracy	Test Accuracy	Test Balanced Accuracy
LogisticRegression	88.84	73.33	89.4	69.11
RandomForestClassifier	86.48	61.0	88.04	54.17
SVC	89.75	73.46	89.95	69.43
GradientBoostingClassifier	95.64	87.3	89.13	62.76
AdaBoostClassifier	90.2	75.83	87.77	63.75
DecisionTreeClassifier	89.47	73.3	85.33	55.26
VotingClassifier	91.38	75.08	88.86	59.06

CONFUSION MATRIX - TRAIN AND TEST for each model...

Key...

Category	Predicted to be False	Predicted to be True
Actually false	True Negative (TN)	False Positive (FP)
Actually true	False Negative (FN)	True Positive (TP)

CONFUSION MATRIX FOR EACH MODEL

Model LogisticRegression Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 885	FP: 28
Actually true	FN: 95	TP: 94

For Training, if '1' predicted, it is correct 77.05% of the time

For Training, if '0' predicted, it is correct 90.31% of the time

Model LogisticRegression TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 309	FP: 11
Actually true	FN: 28	TP: 20

During TEST, if '1' predicted, it is correct 64.52% of the time

During TEST, if '0' predicted, it is correct 91.69% of the time

Model RandomForestClassifier Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 911	FP: 2
Actually true	FN: 147	TP: 42

For Training, if '1' predicted, it is correct 95.45% of the time

For Training, if '0' predicted, it is correct 86.11% of the time

Model RandomForestClassifier TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 320	FP: 0
Actually true	FN: 44	TP: 4

During TEST, if '1' predicted, it is correct 100.0% of the time

During TEST, if '0' predicted, it is correct 87.91% of the time

Model SVC Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 897	FP: 16
Actually true	FN: 97	TP: 92

For Training, if '1' predicted, it is correct 85.19% of the time

For Training, if '0' predicted, it is correct 90.24% of the time

Model SVC TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 311	FP: 9
Actually true	FN: 28	TP: 20

During TEST, if '1' predicted, it is correct 68.97% of the time

During TEST, if '0' predicted, it is correct 91.74% of the time

Model GradientBoostingClassifier Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 913	FP: 0
Actually true	FN: 48	TP: 141

For Training, if '1' predicted, it is correct 100.0% of the time

For Training, if '0' predicted, it is correct 95.01% of the time

Model GradientBoostingClassifier TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 315	FP: 5

Category	Predicted to be False	Predicted to be True
Actually true	FN: 35	TP: 13

During TEST, if '1' predicted, it is correct 72.22% of the time

During TEST, if '0' predicted, it is correct 90.0% of the time

Model AdaBoostClassifier Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 892	FP: 21
Actually true	FN: 87	TP: 102

For Training, if '1' predicted, it is correct 82.93% of the time

For Training, if '0' predicted, it is correct 91.11% of the time

Model AdaBoostClassifier TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 308	FP: 12
Actually true	FN: 33	TP: 15

During TEST, if '1' predicted, it is correct 55.56% of the time

During TEST, if '0' predicted, it is correct 90.32% of the time

Model DecisionTreeClassifier Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 894	FP: 19
Actually true	FN: 97	TP: 92

For Training, if '1' predicted, it is correct 82.88% of the time

For Training, if '0' predicted, it is correct 90.21% of the time

Model DecisionTreeClassifier TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 307	FP: 13
Actually true	FN: 41	TP: 7

During TEST, if '1' predicted, it is correct 35.0% of the time

During TEST, if '0' predicted, it is correct 88.22% of the time

Model VotingClassifier Training CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 912	FP: 1

Category	Predicted to be False	Predicted to be True
Actually true	FN: 94	TP: 95

For Training, if '1' predicted, it is correct 98.96% of the time

For Training, if '0' predicted, it is correct 90.66% of the time

Model VotingClassifier TEST CM

Category	Predicted to be False	Predicted to be True
Actually false	TN: 318	FP: 2
Actually true	FN: 39	TP: 9

During TEST, if '1' predicted, it is correct 81.82% of the time

During TEST, if '0' predicted, it is correct 89.08% of the time

Classification Matrix for each model

Model LogisticRegression Training Classification Report

```
precision    recall  f1-score   support

           0       0.90      0.97      0.94       913
           1       0.77      0.50      0.60       189
```

accuracy			0.89	1102
macro avg	0.84	0.73	0.77	1102
weighted avg	0.88	0.89	0.88	1102

Model LogisticRegression TEST Classification Report

precision	recall	f1-score	support
0	0.92	0.97	0.94	320
1	0.65	0.42	0.51	48
accuracy			0.89	368
macro avg	0.78	0.69	0.72	368
weighted avg	0.88	0.89	0.88	368

Model RandomForestClassifier Training Classification Report

precision	recall	f1-score	support
0	0.86	1.00	0.92	913
1	0.95	0.22	0.36	189
accuracy			0.86	1102
macro avg	0.91	0.61	0.64	1102
weighted avg	0.88	0.86	0.83	1102

Model RandomForestClassifier TEST Classification Report

precision	recall	f1-score	support
0	0.88	1.00	0.94	320
1	1.00	0.08	0.15	48

accuracy			0.88	368
macro avg	0.94	0.54	0.54	368
weighted avg	0.89	0.88	0.83	368

Model SVC Training Classification Report

precision	recall	f1-score	support
0	0.90	0.98	0.94	913
1	0.85	0.49	0.62	189
accuracy			0.90	1102
macro avg	0.88	0.73	0.78	1102
weighted avg	0.89	0.90	0.89	1102

Model SVC TEST Classification Report

precision	recall	f1-score	support
0	0.92	0.97	0.94	320
1	0.69	0.42	0.52	48
accuracy			0.90	368
macro avg	0.80	0.69	0.73	368
weighted avg	0.89	0.90	0.89	368

Model GradientBoostingClassifier Training Classification Report

precision	recall	f1-score	support
0	0.95	1.00	0.97	913
1	1.00	0.75	0.85	189

accuracy			0.96	1102
macro avg	0.98	0.87	0.91	1102
weighted avg	0.96	0.96	0.95	1102

Model GradientBoostingClassifier TEST Classification Report

precision	recall	f1-score	support
0	0.90	0.98	0.94	320
1	0.72	0.27	0.39	48
accuracy			0.89	368
macro avg	0.81	0.63	0.67	368
weighted avg	0.88	0.89	0.87	368

Model AdaBoostClassifier Training Classification Report

precision	recall	f1-score	support
0	0.91	0.98	0.94	913
1	0.83	0.54	0.65	189
accuracy			0.90	1102
macro avg	0.87	0.76	0.80	1102
weighted avg	0.90	0.90	0.89	1102

Model AdaBoostClassifier TEST Classification Report

precision	recall	f1-score	support
0	0.90	0.96	0.93	320
1	0.56	0.31	0.40	48
accuracy			0.88	368

```
   macro avg       0.73      0.64      0.67       368
weighted avg       0.86      0.88      0.86       368
```

Model DecisionTreeClassifier Training Classification Report

```
precision    recall  f1-score   support

           0       0.90      0.98      0.94       913
           1       0.83      0.49      0.61       189

    accuracy                           0.89      1102
   macro avg       0.87      0.73      0.78      1102
weighted avg       0.89      0.89      0.88      1102
```

Model DecisionTreeClassifier TEST Classification Report

```
precision    recall  f1-score   support

           0       0.88      0.96      0.92       320
           1       0.35      0.15      0.21        48

    accuracy                           0.85       368
   macro avg       0.62      0.55      0.56       368
weighted avg       0.81      0.85      0.83       368
```

Model VotingClassifier Training Classification Report

```
precision    recall  f1-score   support

           0       0.91      1.00      0.95       913
           1       0.99      0.50      0.67       189

    accuracy                           0.91      1102
```

macro avg	0.95	0.75	0.81	1102
weighted avg	0.92	0.91	0.90	1102

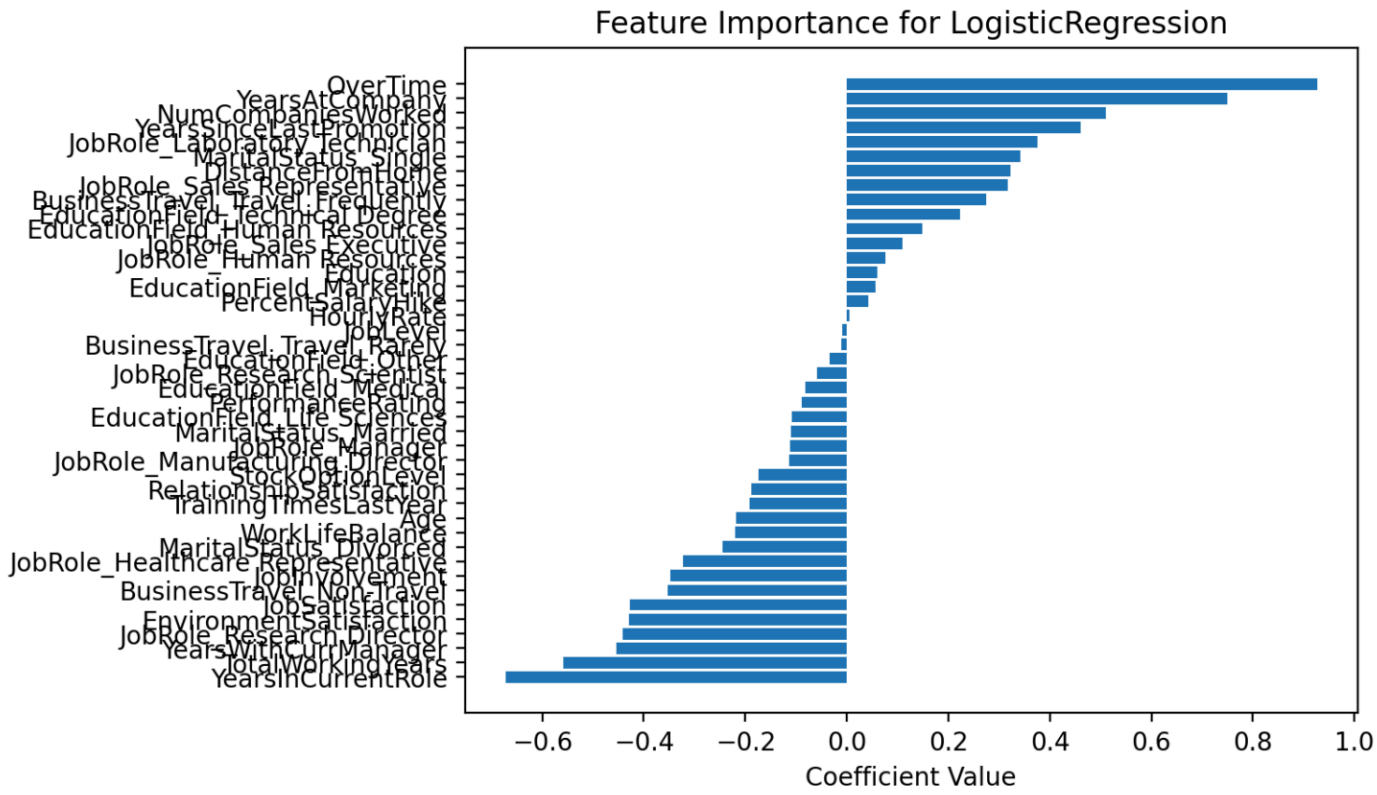
Model VotingClassifier TEST Classification Report

	precision	recall	f1-score	support
0	0.89	0.99	0.94	320
1	0.82	0.19	0.31	48
accuracy			0.89	368
macro avg	0.85	0.59	0.62	368
weighted avg	0.88	0.89	0.86	368

Feature Importance for each model

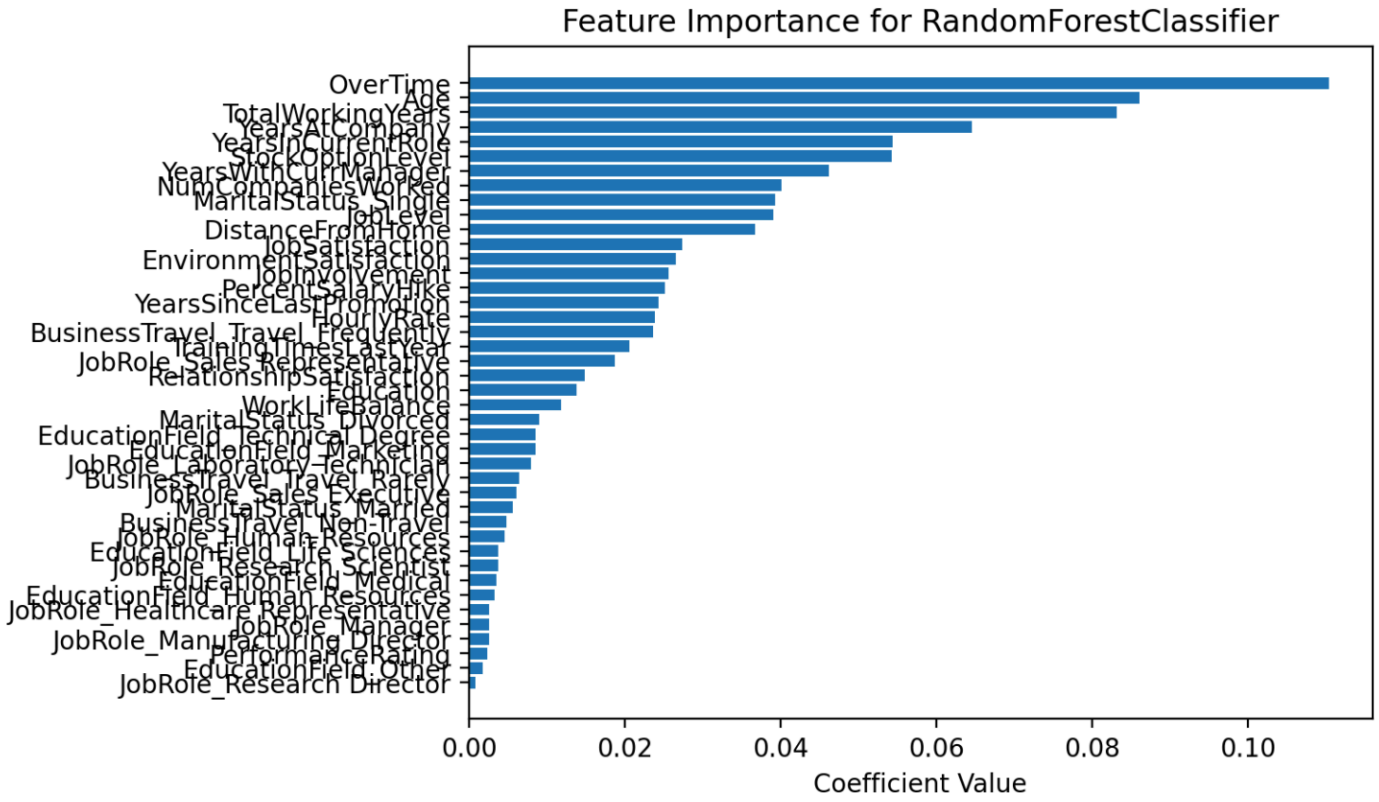
Model LogisticRegression Feature Importance

	feature_names	feature_importance
18	MaritalStatus_Divorced	-0.2455
0	JobRole_Healthcare Representative	-0.3222
35	JobInvolvement	-0.3481
15	BusinessTravel_Non-Travel	-0.3529
33	JobSatisfaction	-0.4278
37	EnvironmentSatisfaction	-0.4291
5	JobRole_Research Director	-0.441
40	YearsWithCurrManager	-0.4534
26	TotalWorkingYears	-0.5585
22	YearsInCurrentRole	-0.6724



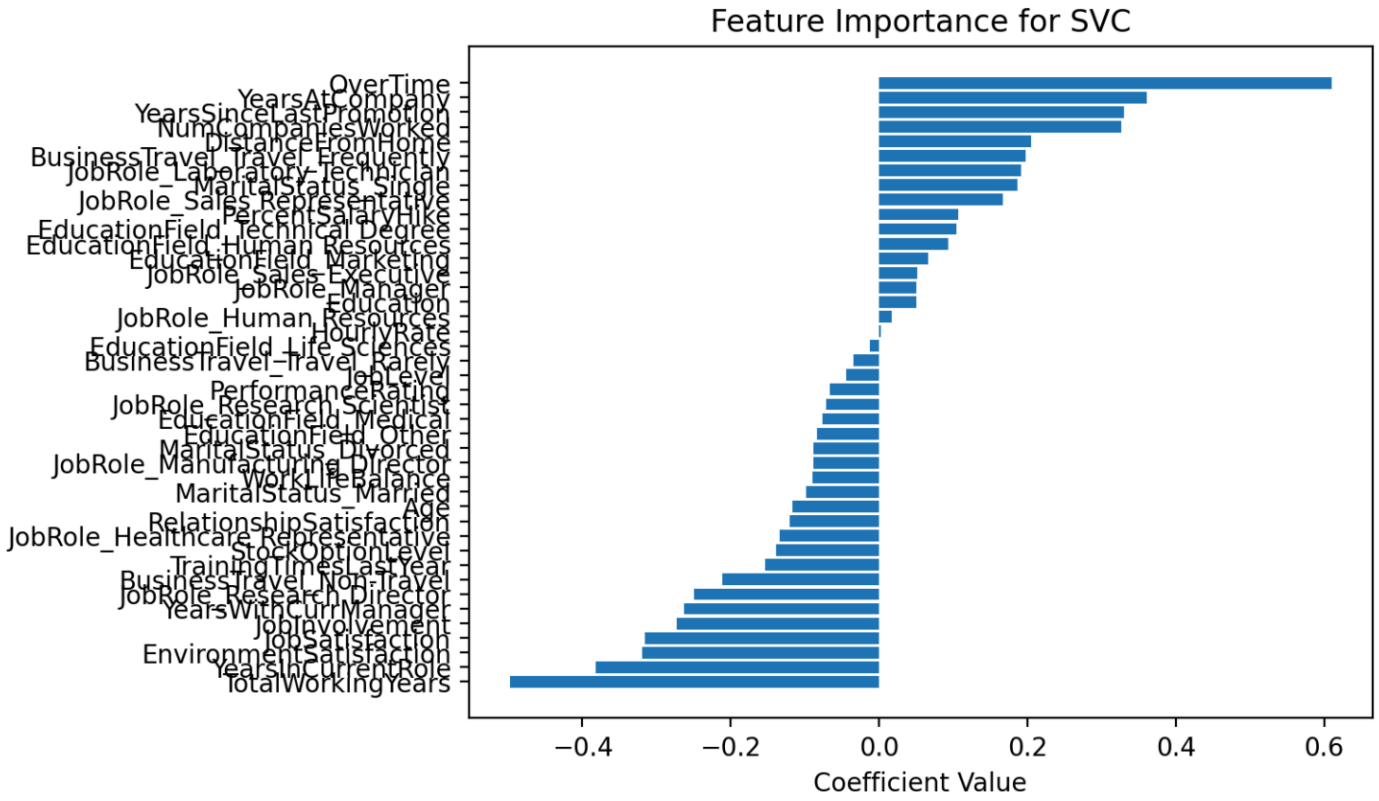
Model RandomForestClassifier Feature Importance

	feature_names	feature_importance
41	OverTime	0.1104
21	Age	0.0861
26	TotalWorkingYears	0.0832
23	YearsAtCompany	0.0646
22	YearsInCurrentRole	0.0543
27	StockOptionLevel	0.0543
40	YearsWithCurrManager	0.0462
32	NumCompaniesWorked	0.0401
20	MaritalStatus_Single	0.0393
34	JobLevel	0.0391



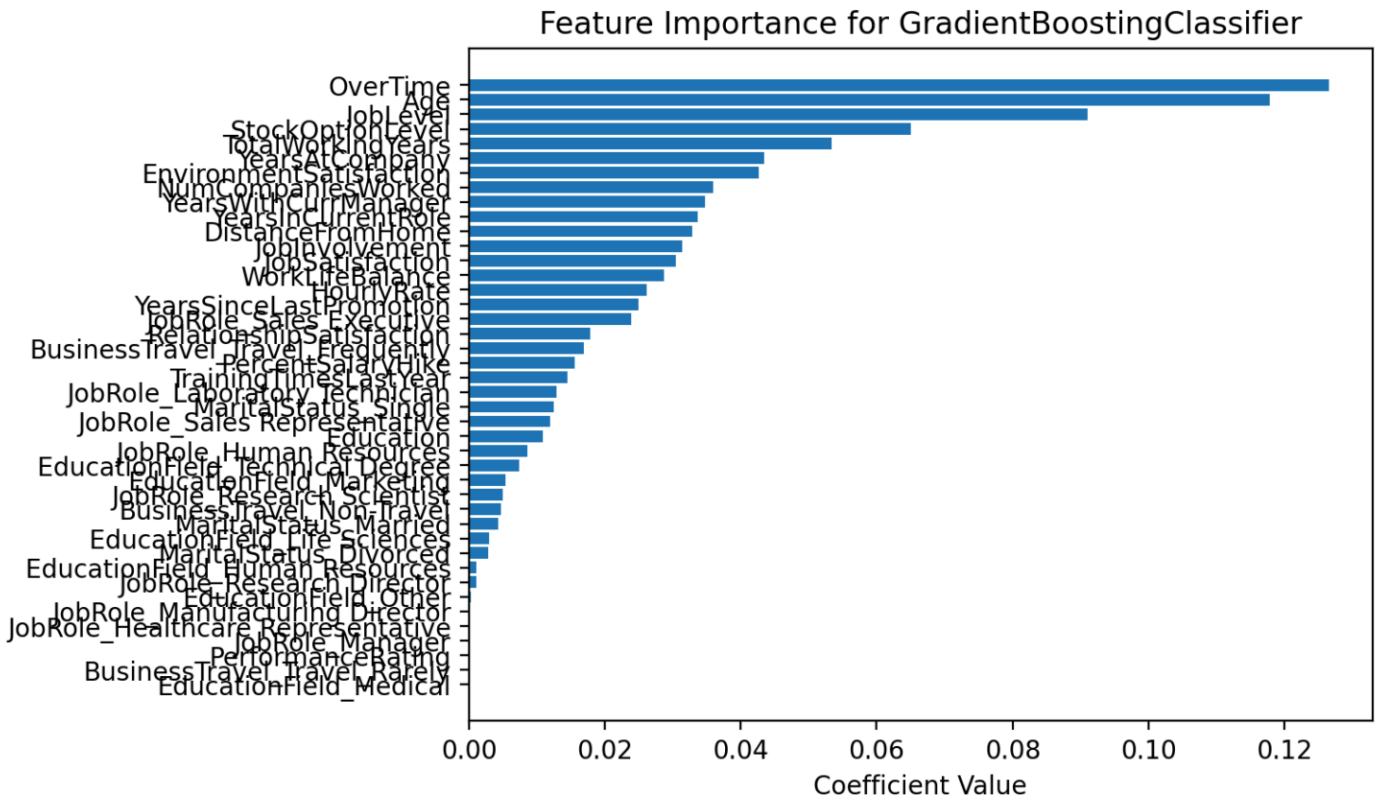
Model SVC Feature Importance

	feature_names	feature_importance
41	OverTime	0.6098
23	YearsAtCompany	0.3607
31	YearsSinceLastPromotion	0.3305
32	NumCompaniesWorked	0.3262
39	DistanceFromHome	0.2053
16	BusinessTravel_Travel_Frequently	0.1971
2	JobRole_Laboratory Technician	0.1918
20	MaritalStatus_Single	0.1866
8	JobRole_Sales Representative	0.1666
30	PercentSalaryHike	0.1066



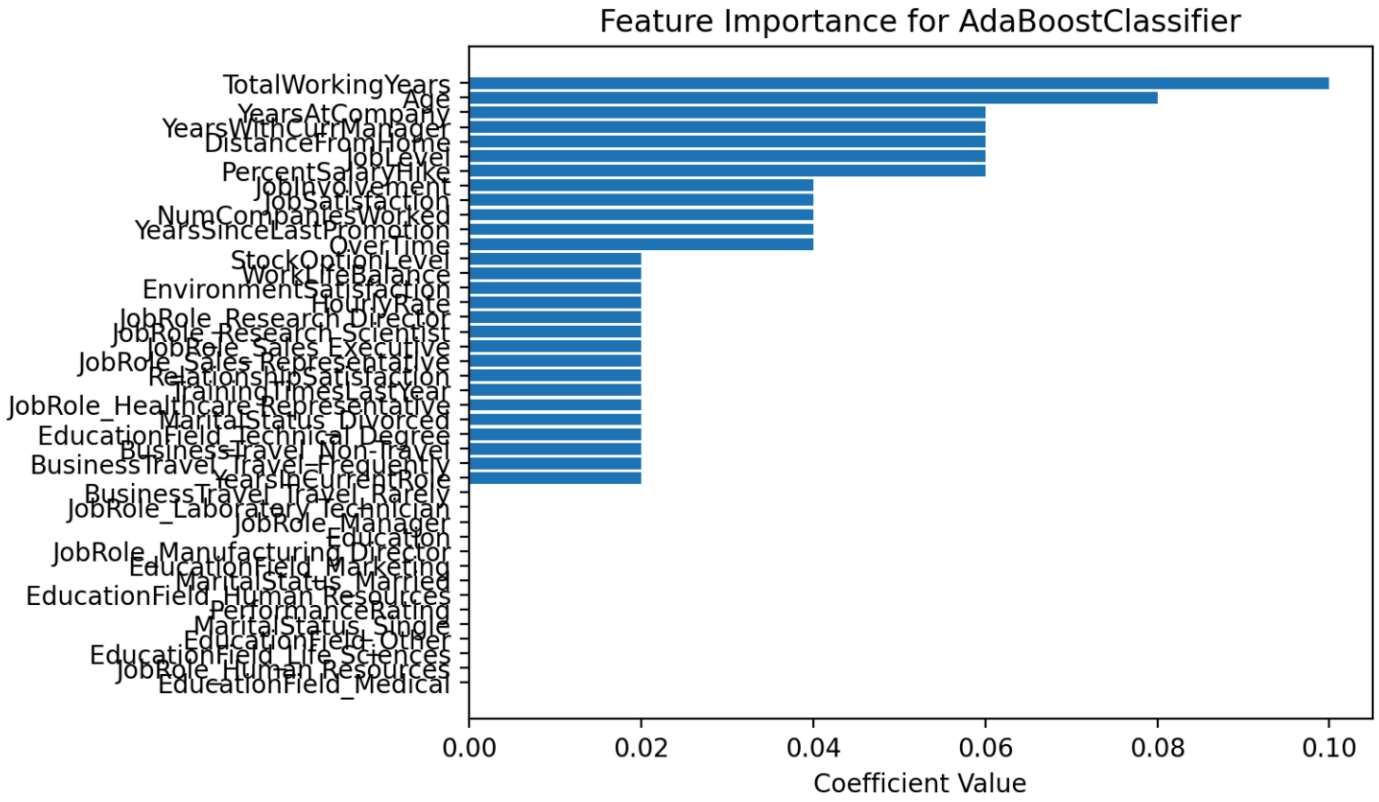
Model GradientBoostingClassifier Feature Importance

	feature_names	feature_importance
28	RelationshipSatisfaction	0.0179
16	BusinessTravel_Travel_Fre	0.0169
30	PercentSalaryHike	0.0156
25	TrainingTimesLastYear	0.0144
2	JobRole_Laboratory Techı	0.0129
20	MaritalStatus_Single	0.0125
8	JobRole_Sales Representa	0.012
38	Education	0.0109
1	JobRole_Human Resource	0.0086
14	EducationField_Technical	0.0074
11	EducationField_Marketing	0.0054



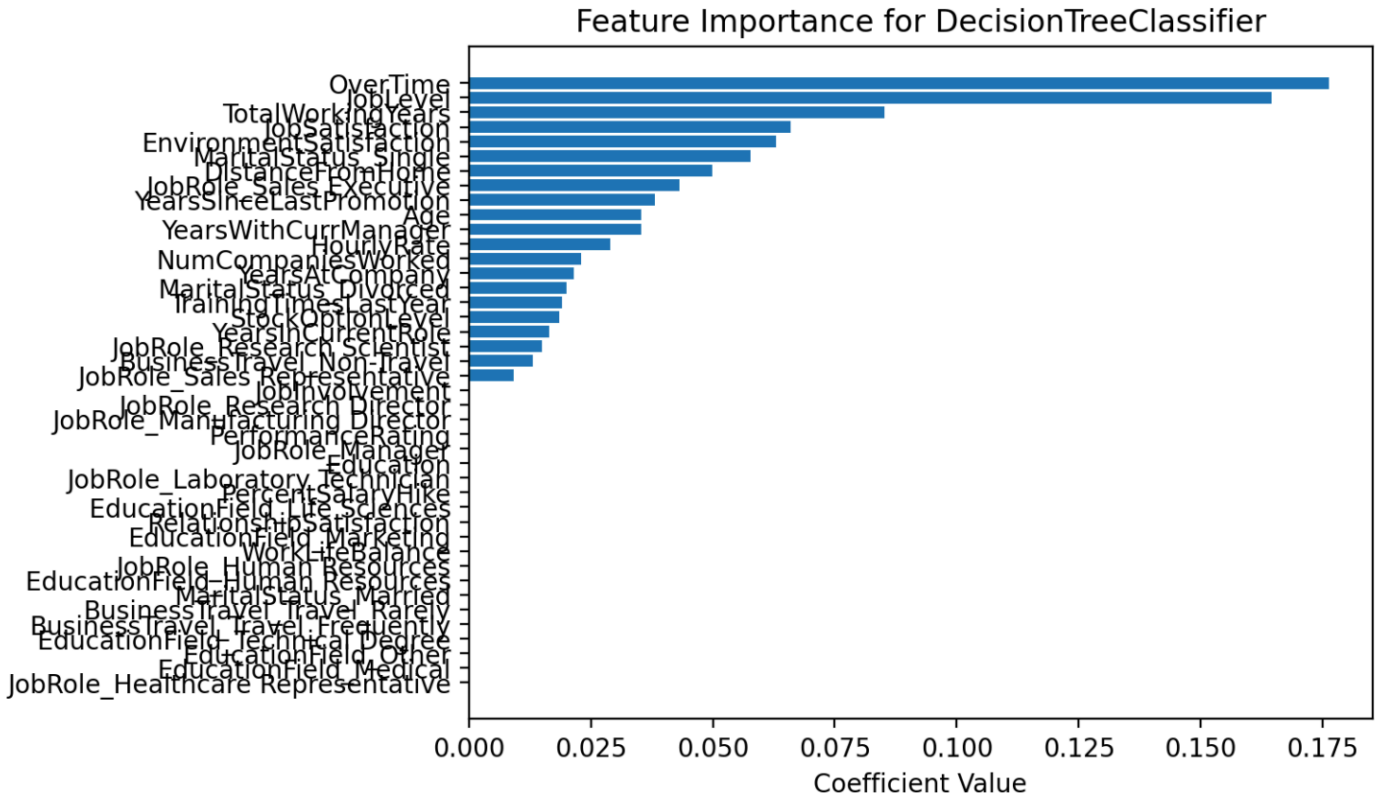
Model AdaBoostClassifier Feature Importance

	feature_names	feature_importance
26	TotalWorkingYears	0.1
21	Age	0.08
23	YearsAtCompany	0.06
40	YearsWithCurrManager	0.06
39	DistanceFromHome	0.06
34	JobLevel	0.06
30	PercentSalaryHike	0.06
35	JobInvolvement	0.04
33	JobSatisfaction	0.04
32	NumCompaniesWorked	0.04



Model DecisionTreeClassifier Feature Importance

	feature_names	feature_importance
41	OverTime	0.1764
34	JobLevel	0.1646
26	TotalWorkingYears	0.0851
33	JobSatisfaction	0.066
37	EnvironmentSatisfaction	0.0629
20	MaritalStatus_Single	0.0578
39	DistanceFromHome	0.0499
7	JobRole_Sales Executive	0.0432
31	YearsSinceLastPromotion	0.0382
21	Age	0.0354



Analysis Complete