Predicting Salary Based on Online Job Posting

MACHINE LEARNING

El Futuro...

Presented by **Kuntal Adak**

linkedin.com/in/adakkuntal

WE ARE HIRING !

The Problem Statement...

IN THIS PROJECT, I WILL BE USING MACHINE LEARNING MODELS SUCH AS LINEAR REGRESSION AND RANDOM FOREST ON THE DATASET TO PREDICT SALARIES OF SCIENTIST, ANALYST AND ENGINEER.

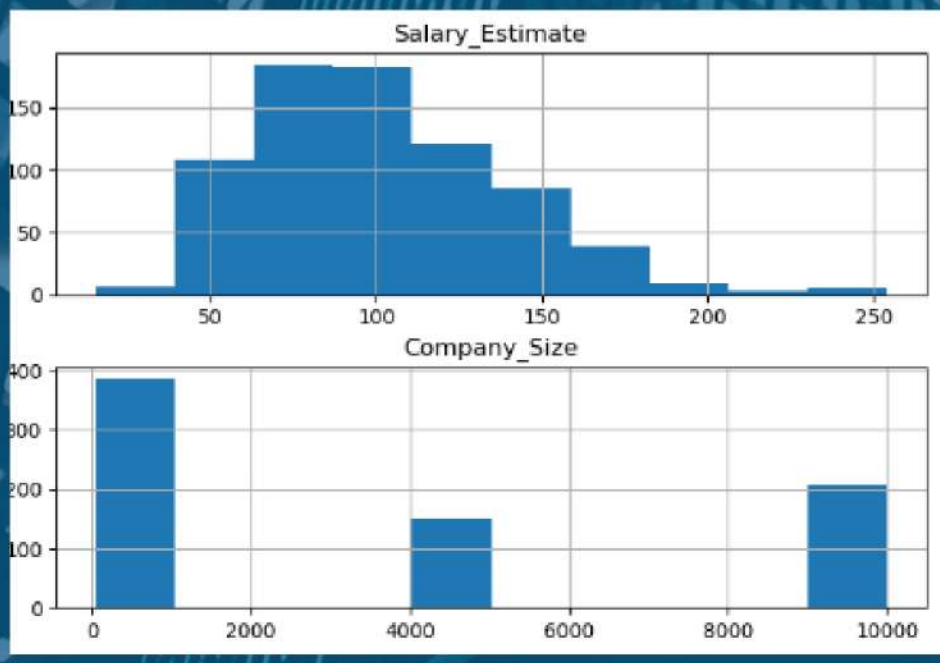
Easy Apply

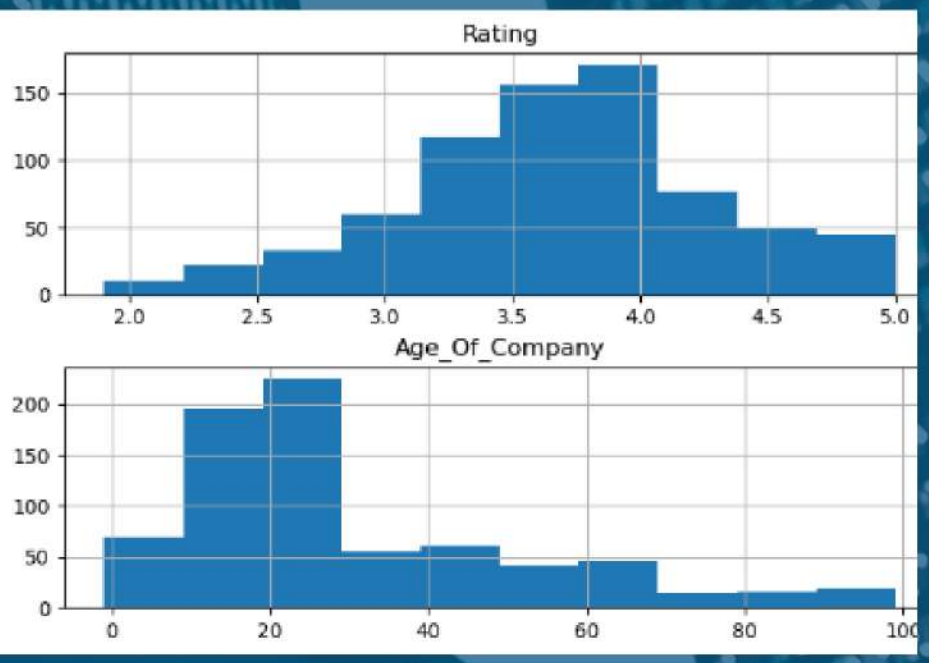
FEATURE ENGINEERING

Dataset Columns	Newly Created Columns
Job Title	Title (Scientist, Analyst, Engineer) Position (Senior, Junior)
Job Description	Degree_Holder Experienced Communication Visualization Machine_learning Python Spark AWS Excel SQL
Founded	Age_Of_Company
Competitors	Total_Competitors
Salary Estimate	Per_Hour_Salary Employer_Salary Average Salary (Target Variable)

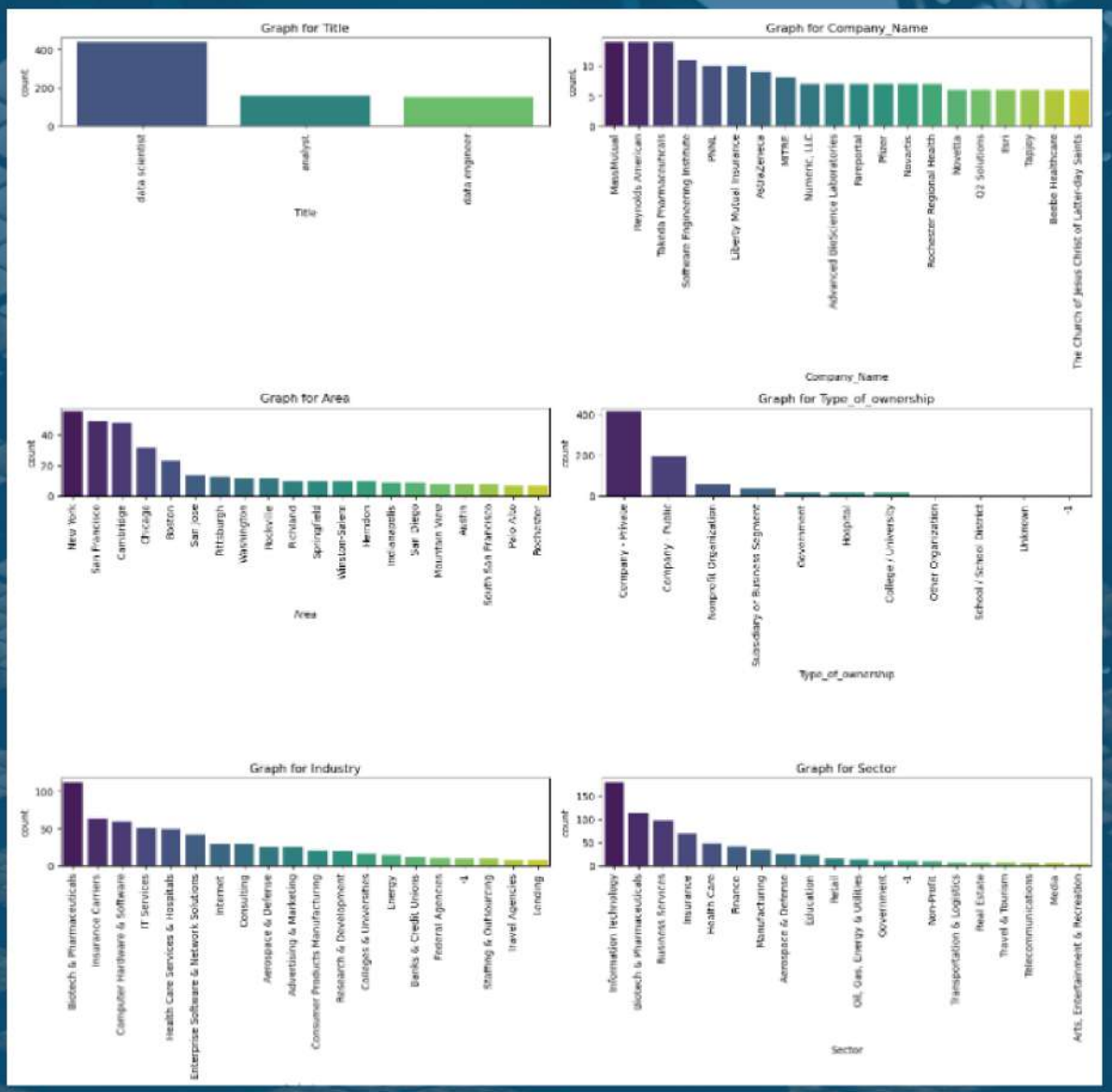
Find The Perfect Job That You Deserve.

EDA





CORRELATION MATRIX



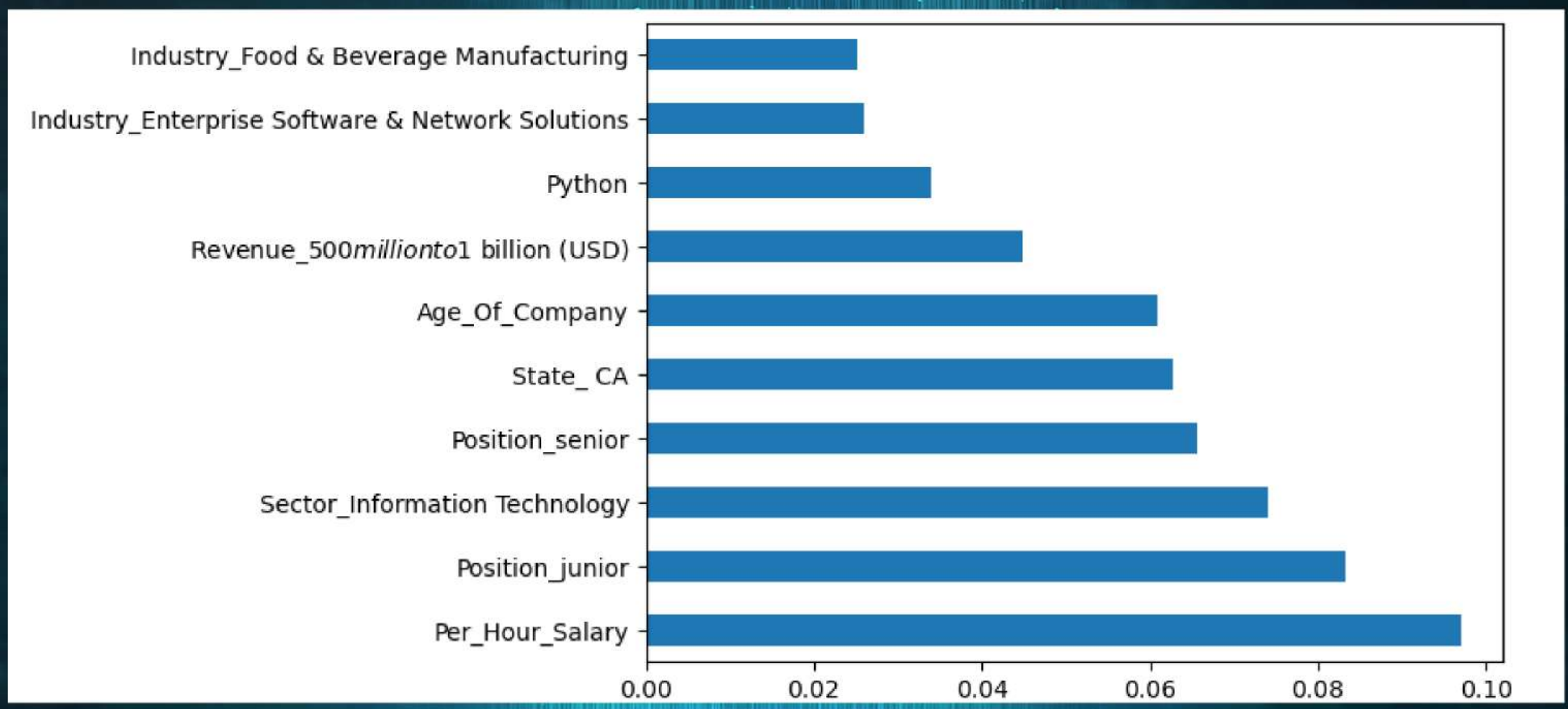
MODEL SELECTION

```
LinearRegression()
R2 Value: 1.0

##### Model Validation and Accuracy Calculations ##########
   Salary_Estimate  PredictedSalary_Estimate        APE
0              172                      90.0  47.674419
1               97                      61.0  37.113402
2               76                      83.0   9.210526
3               86                     110.0  27.906977
4               73                      73.0   0.000000
Mean Accuracy on test data: 88.92196895181061
Median Accuracy on test data: 100.0

Accuracy values for 10-fold Cross Validation:
 [ 96.85761771  88.09835581  89.66390248  79.94367674  93.99999872
 100.          89.84583679  97.1078759   92.34849579  95.1093546 ]

Final Average Accuracy of the model: 92.3
```

```
RandomForestRegressor(criterion='friedman_mse', max_depth=10)
R2 Value: 0.9119491987263546

##### Model Validation and Accuracy Calculations ##########
   Salary_Estimate  PredictedSalary_Estimate
0              172                     100.0
1               97                      79.0
2               76                      94.0
3               86                     108.0
4               73                      91.0
Mean Accuracy on test data: 85.00794312110538
Median Accuracy on test data: 89.18831168831169

Accuracy values for 10-fold Cross Validation:
 [88.39521579 82.14371483 84.04251282 76.56763088 90.48911976 90.60949175
 85.62196357 87.89674915 87.13890468 89.69391967]

Final Average Accuracy of the model: 86.26
```

BASED ON THE RESULT - LINEAR REGRESSION IS OVER FITTING SO WE CHOOSE RANDOM FOREST MODEL

TOP PREDICTORS



THIS GRAPH SHOWS HOW MUCH FEATURE ENGINEERING IS IMPORTANT IN MACHINE LEARNING


Thank You
For Viewing My Presentation
HAVE A GOOD DAY AHEAD
Kuntal Adak
LINKEDIN.COM/IN/ADAKKUNTAL