Income Level Prediction Models

Sample Constraint

Who We Are!

The Dominatricies specialize in finding the sweet spot on any model while being constrained.

“Art lives from constraints and dies from freedom”

—Leonardo da Vinci

Good, Fast or **Cheap**

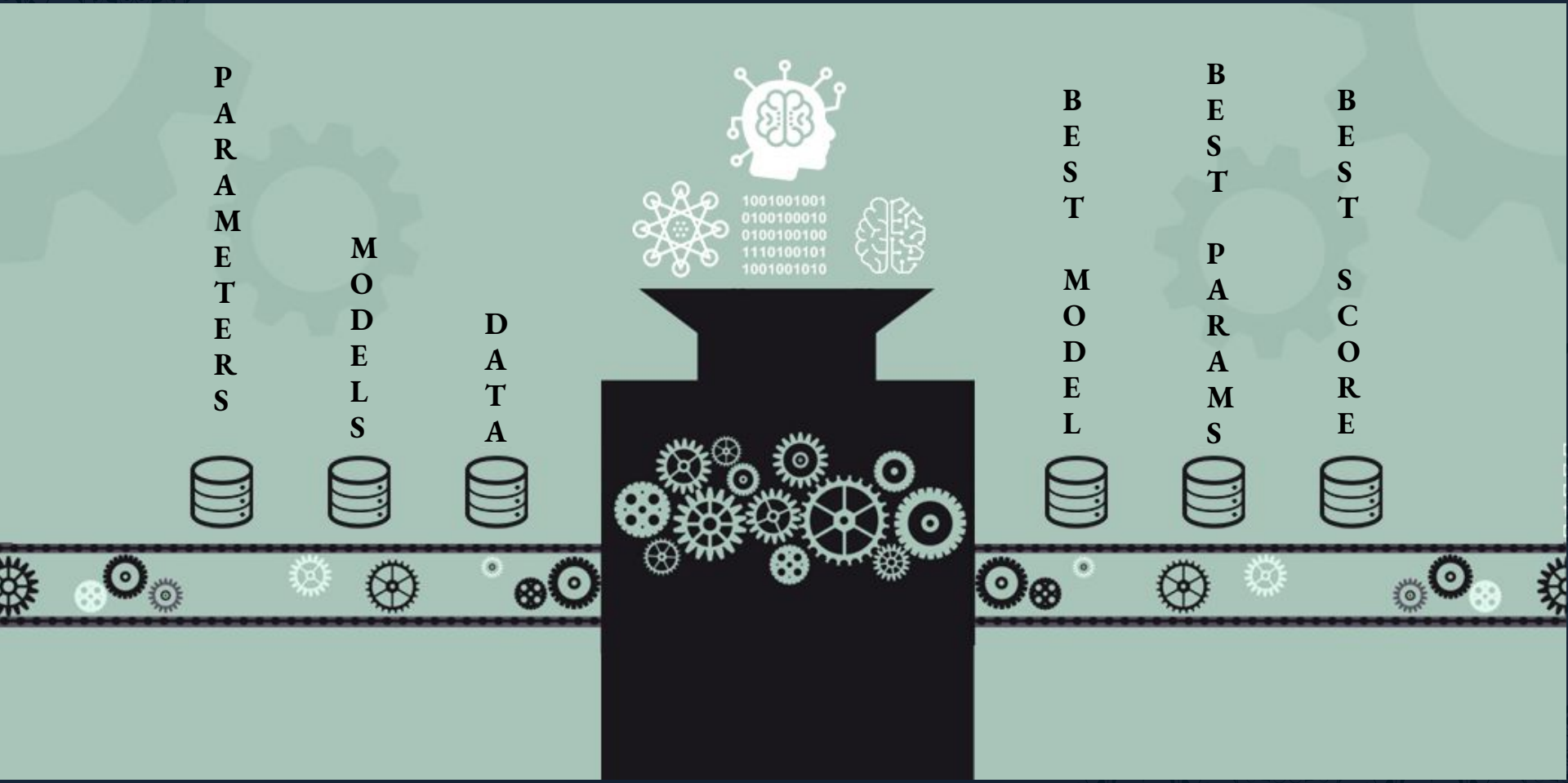
Our model with ⅕ the data allows us to cheaply attain accuracies similar to those with larger data sets.

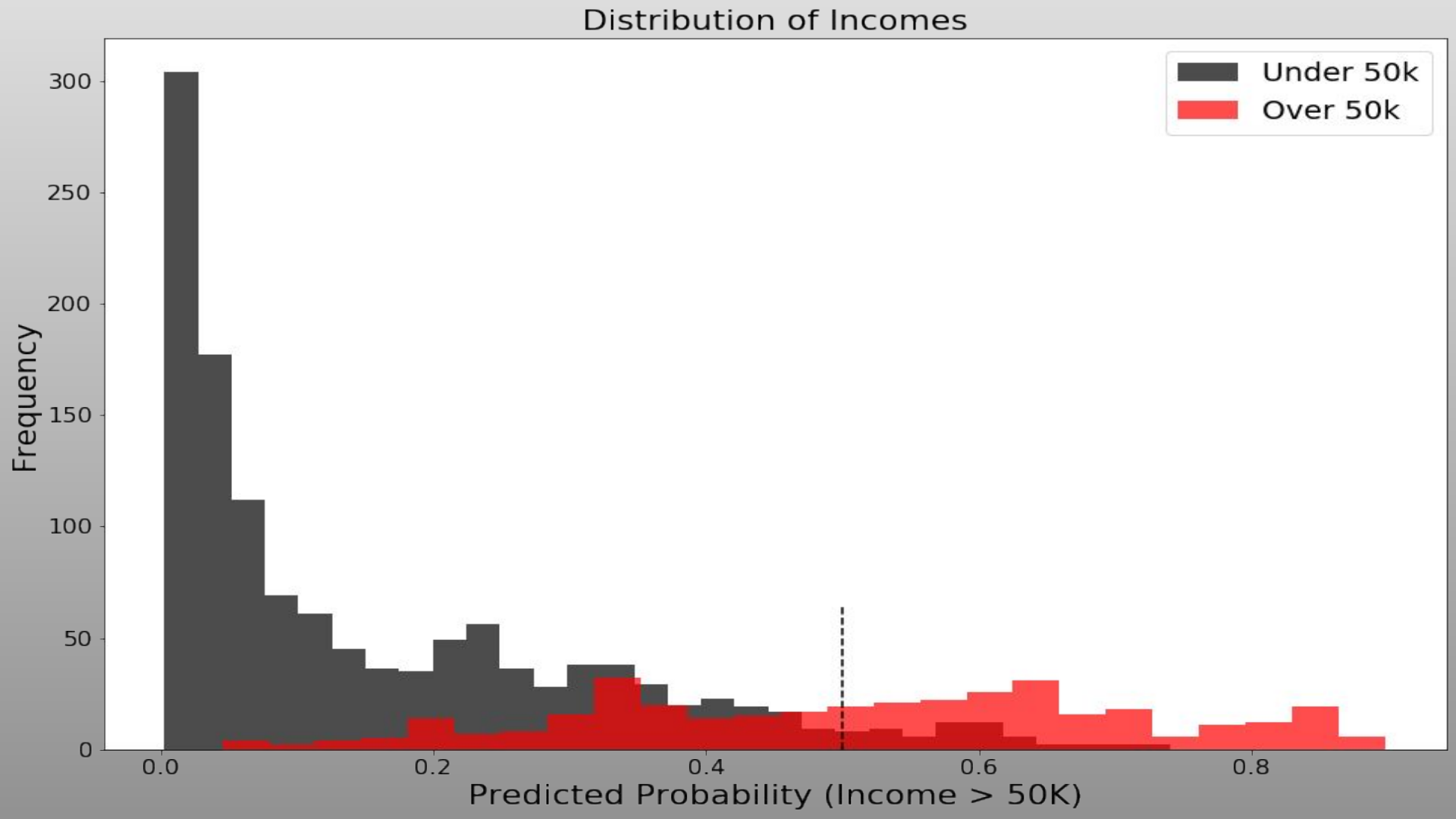
Bootstrapping

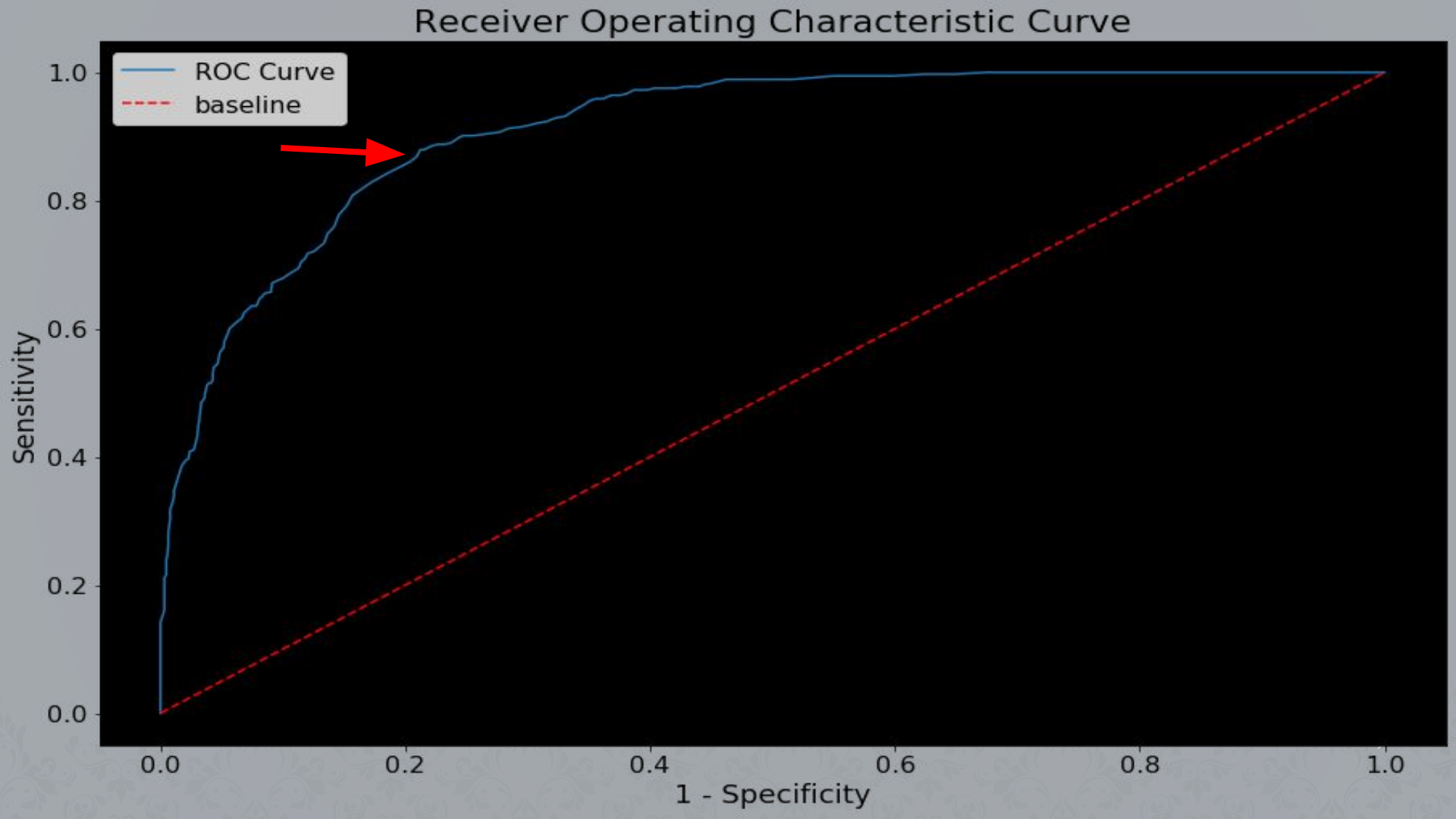
- Random Sampling of current data w/ replacement
- Increase data package to desired size
- Merge new data with existing model

We now have a dataset with over 32,000 rows created with our original 6500.

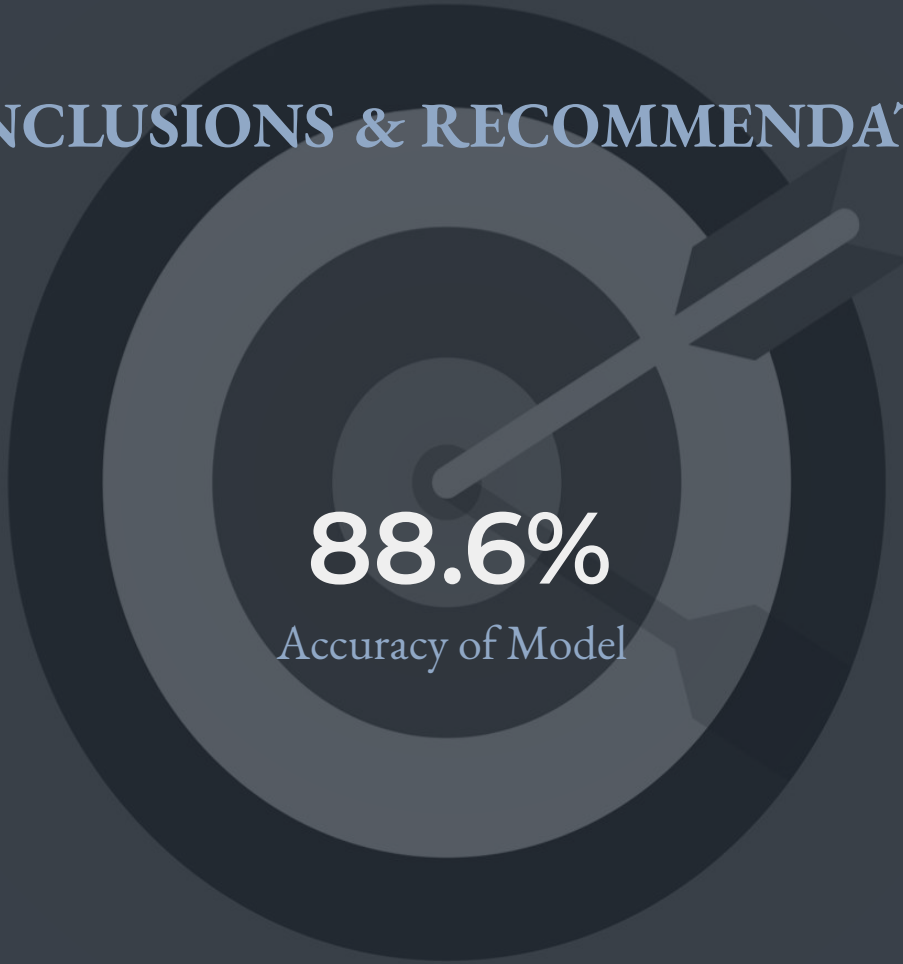
Wage Correlation


PARAMETERS
MODELS
DATA
1001001001
0100100010
0100100100
1110100101
1001001010
BEST MODEL
BEST PARAMS
BEST SCORE


Distribution of Incomes
Under 50k
Over 50k
Frequency
300
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
Predicted Probability (Income > 50K)


Receiver Operating Characteristic Curve
ROC Curve
baseline
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

CONCLUSIONS & RECOMMENDATIONS



88.6%

Accuracy of Model

Thank You!

Does anyone have any questions?

Anshul Tiwari Clint Hoke Daniel Gurzi