Making Money with Lending Club and Data Science

Subramanian Iyer: Project McNulty

Lending Club

-Peer to Peer Lending Platform

-Cuts out bank fees (mostly)

-Interest rates up to 28%

-In theory, great! Let’s make lots of money.

PROBLEM: CHARGE OFF

-Loans with 28% Interest have a default rate of about 50%.

-Works out to a -36% rate of return.

-Can we predict which loans default?

Lending Club Data Decisions

-2013 through 2017q3

-Restrict to data visible from Lending Club's Investor Interface

-Finally: 28 features, 24 numeric, 58 total, with (dummified) categorical

Benchmark: Meet Deepak the monkey

‘What if I gave Deepak the monkey all my money?’

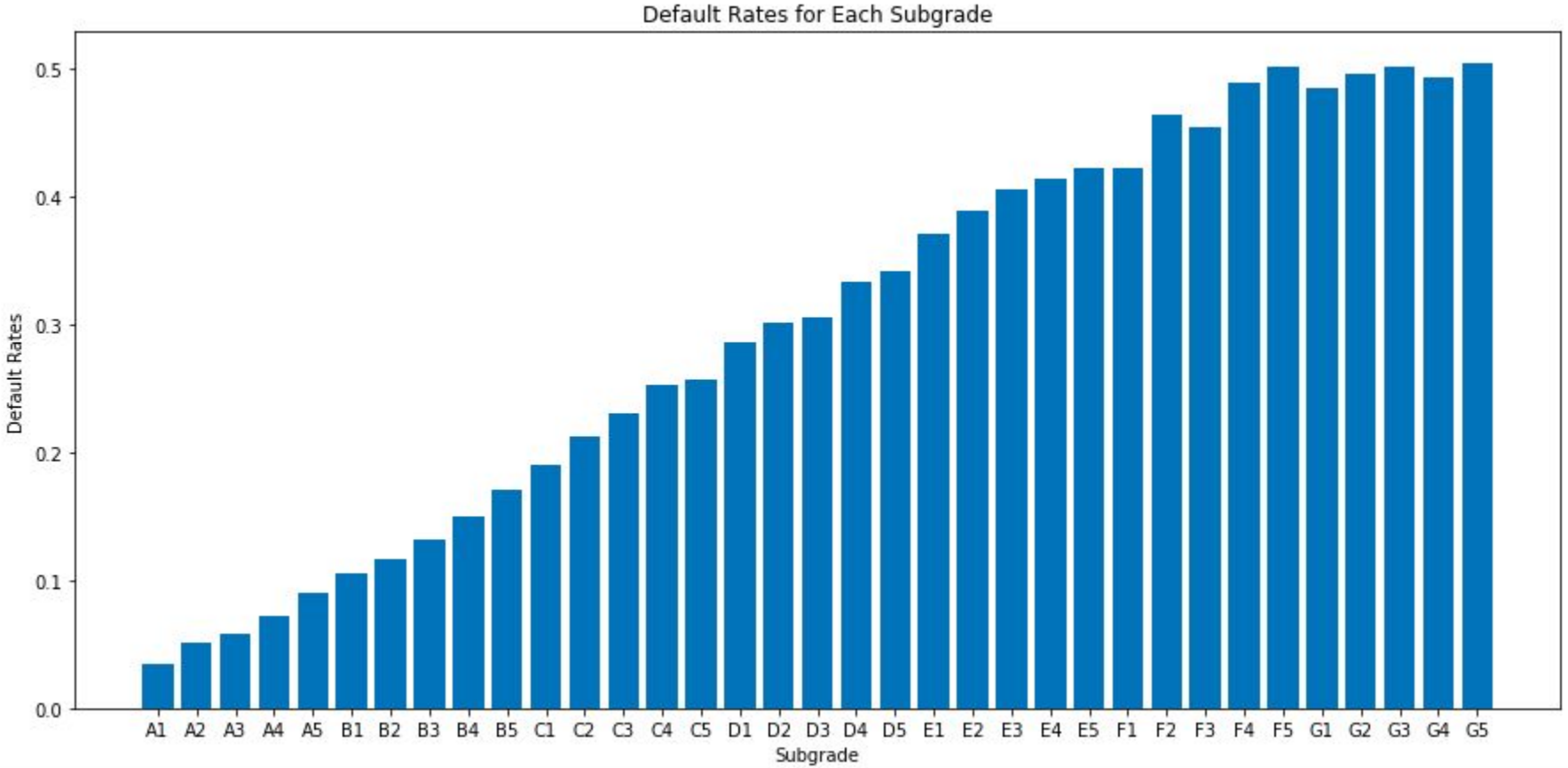
-Deepak tried:

Bagging: Rate of return: -8.5%

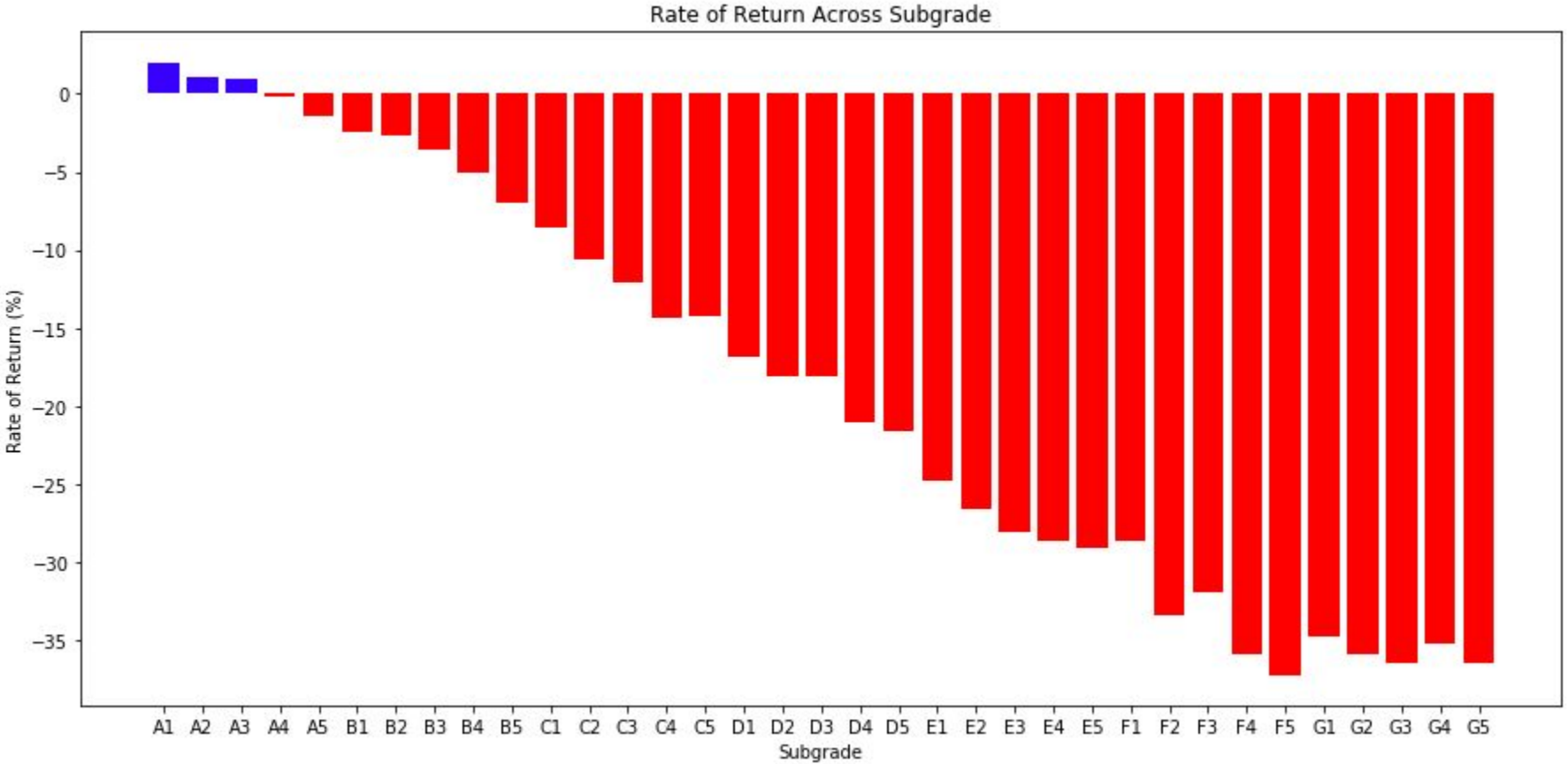
Logistic Regression: Rate of return: -10.2%

Random Forest: -8.8%

Default Rates by Subgrade



Average Return by Subgrade



Process:

On each subgrade:

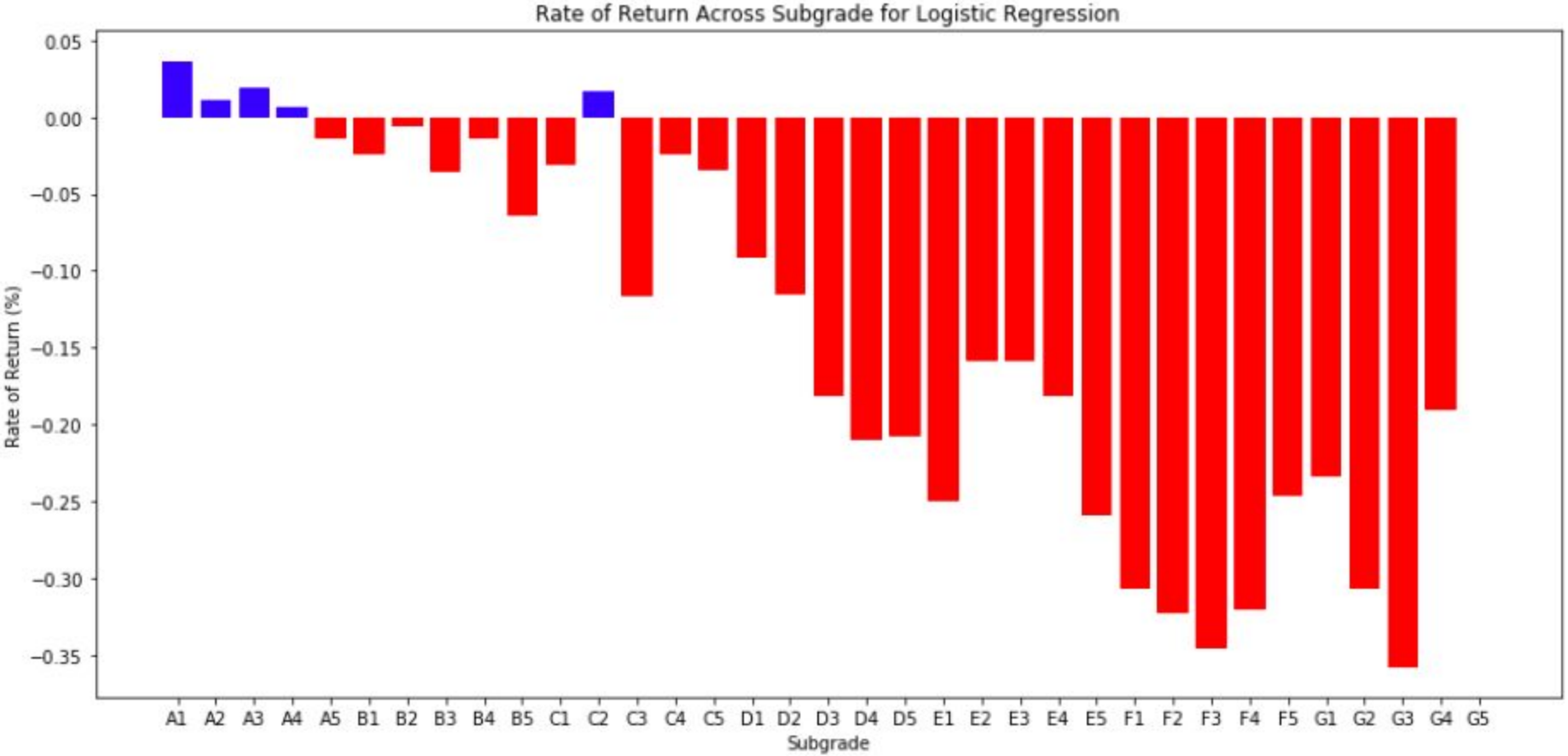
Run Random Forest, Grid Search over:

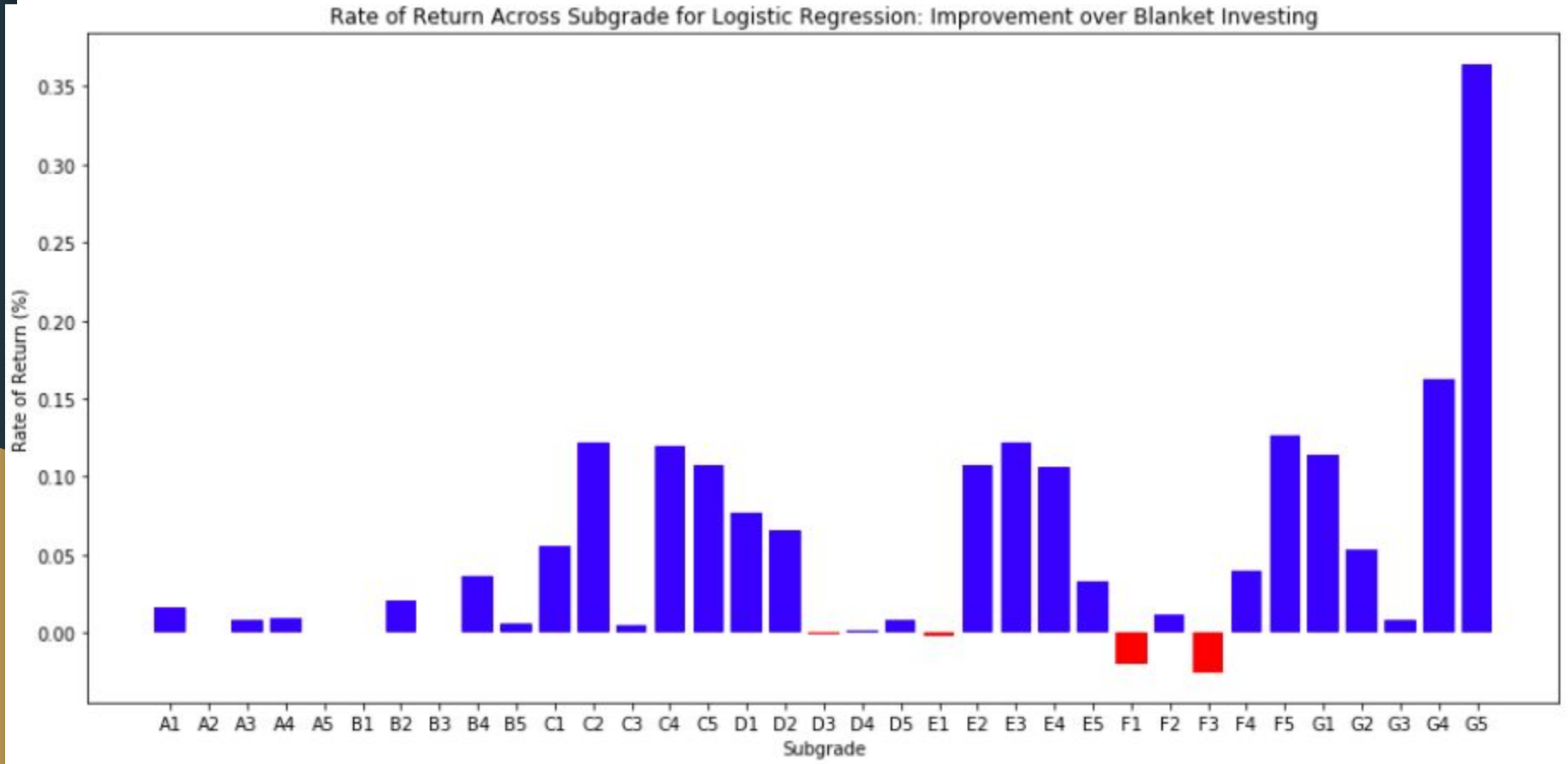
Min sample split, N estimators, max features

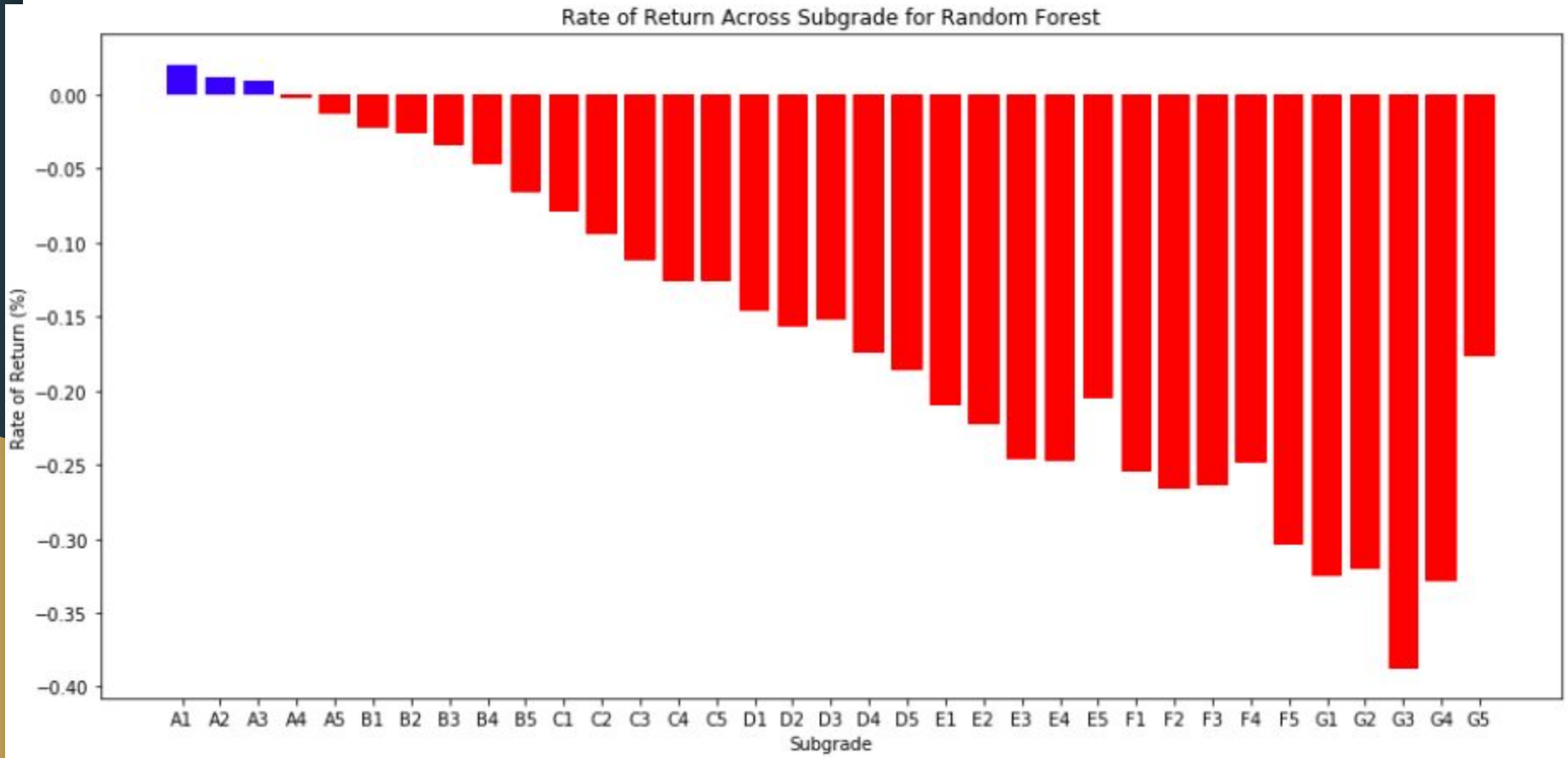
Run Logistic Regression, Grid Search over:

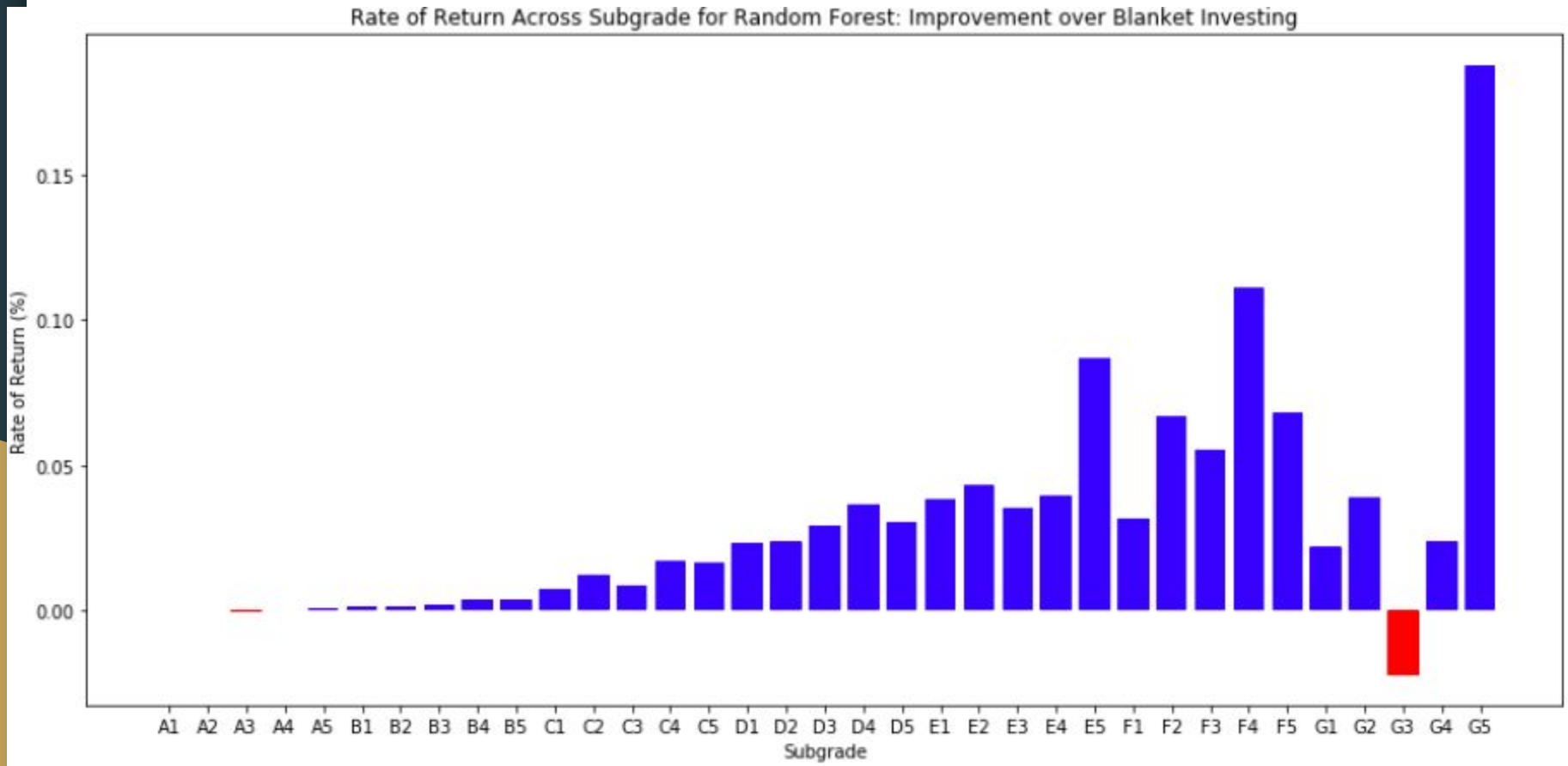
Regularization C

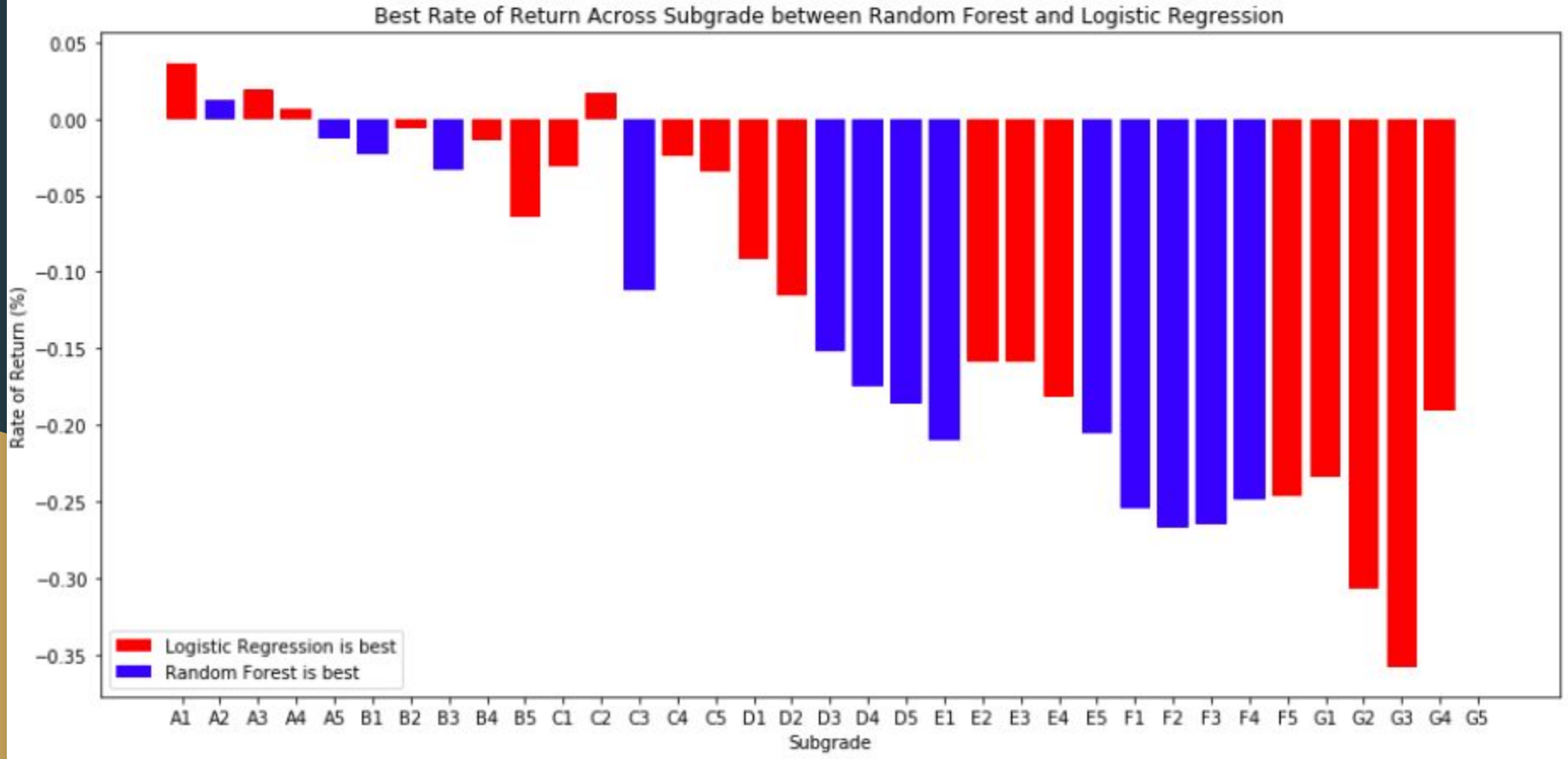
Custom Metric used to judge models: Rate of Return

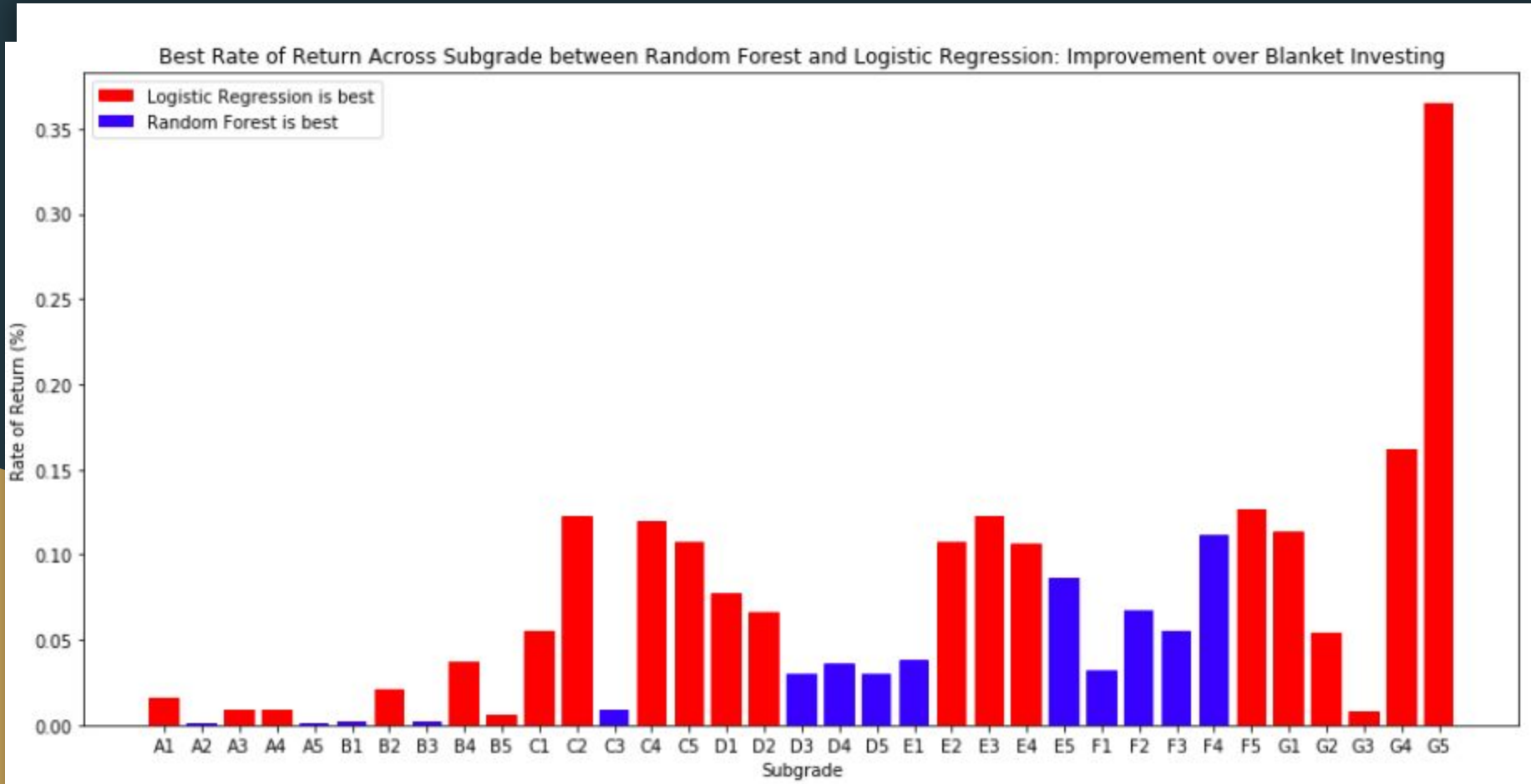

Rate of Return Across Subgrade for Logistic Regression
Rate of Return (%)
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
−0.35
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Subgrade


Rate of Return Across Subgrade for Logistic Regression: Improvement over Blanket Investing
Rate of Return (%)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Subgrade


Rate of Return Across Subgrade for Random Forest
Rate of Return (%)
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
−0.35
−0.40
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Subgrade


Rate of Return Across Subgrade for Random Forest: Improvement over Blanket Investing
Rate of Return (%)
0.15
0.10
0.05
0.00
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Subgrade


Best Rate of Return Across Subgrade between Random Forest and Logistic Regression
Rate of Return (%)
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
−0.35
Logistic Regression is best
Random Forest is best
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Subgrade


Best Rate of Return Across Subgrade between Random Forest and Logistic Regression: Improvement over Blanket Investing
Logistic Regression is best
Random Forest is best
Rate of Return (%)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Subgrade

Final Numbers

Blanket Investing over Everything: -10.2%

Deepak the monkey: -8.5%

Rishabh: 3.64%

Vanguard Total Stock Market Index Fund: 15.52% (avg over 5 years)

Special Thanks

-Laura Chen

-Chris Giler

-Brian Kim

-Jon Kislin

-Malik Salam

-Brad Solomon

-Loren Chen

-Julia Lintern

-David Oxnard

-Vinny Senguttuvan

Questions

Grid Search Parameters

Random Forest:

- Number of trees: 8-23
- Features: 7-34
- Min Samples Split:4-13

Logistic Regression:

C: 10^[1.86,1.87,1.88,1.89,1.90]