Income estimation

Based on skills

Can we calculate salaries for jobs in any industry based only on skills and location?

Datasets

Salary data from Bureau of Labor Statistics:

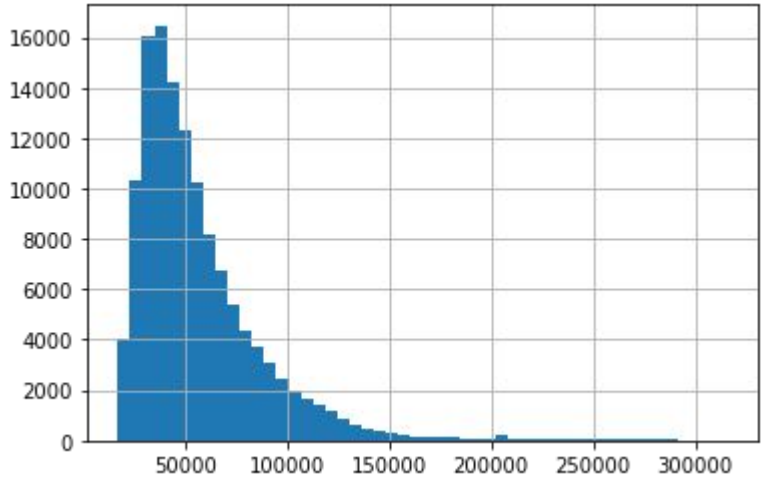
- Occupation
- State
- Annual Salary

Skills data from Onet

There are two scales, impotence and level

For this exercise we choose importance with a scale of 1-5

Target feature



	Annual	Hourly
Mean	57,448	27.34
Min	17,300	8.32
Max	315830	151.84

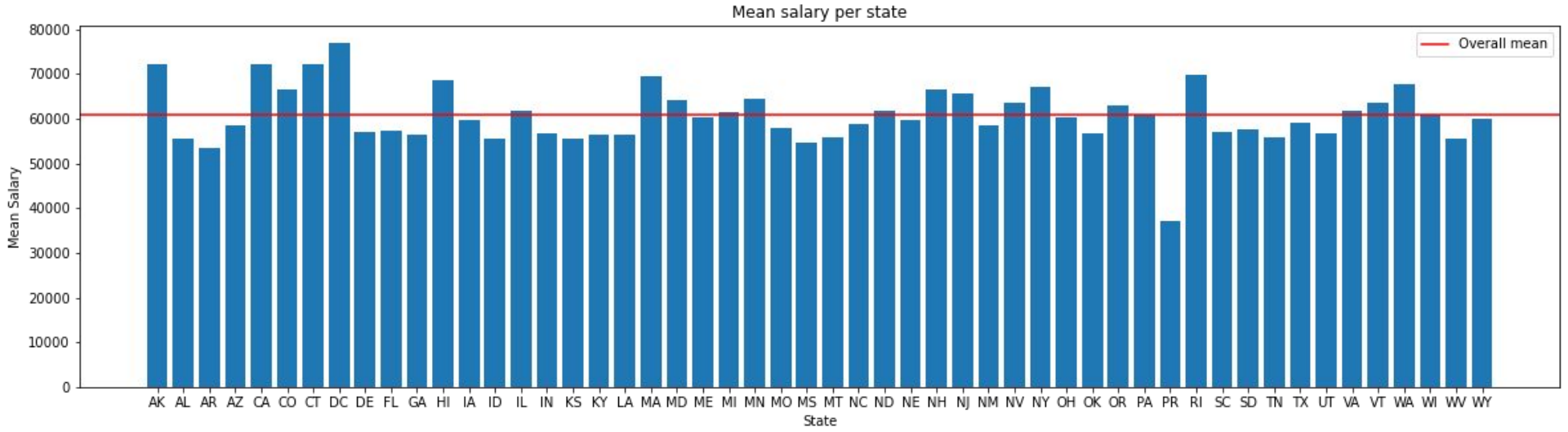
Features

Features

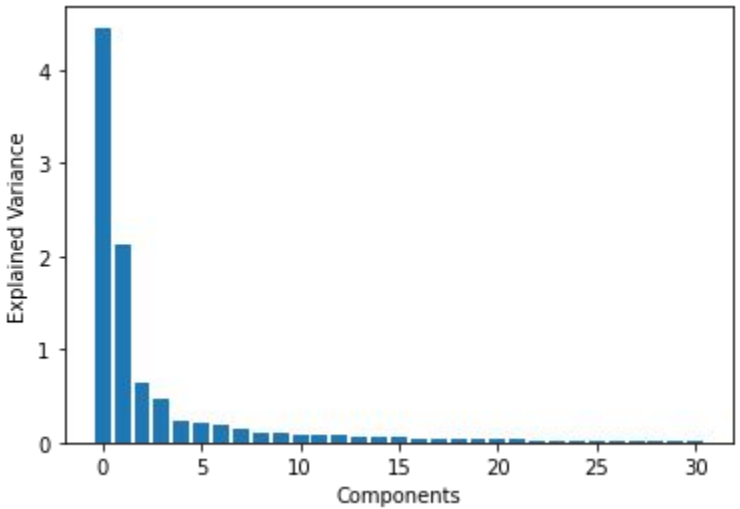
states

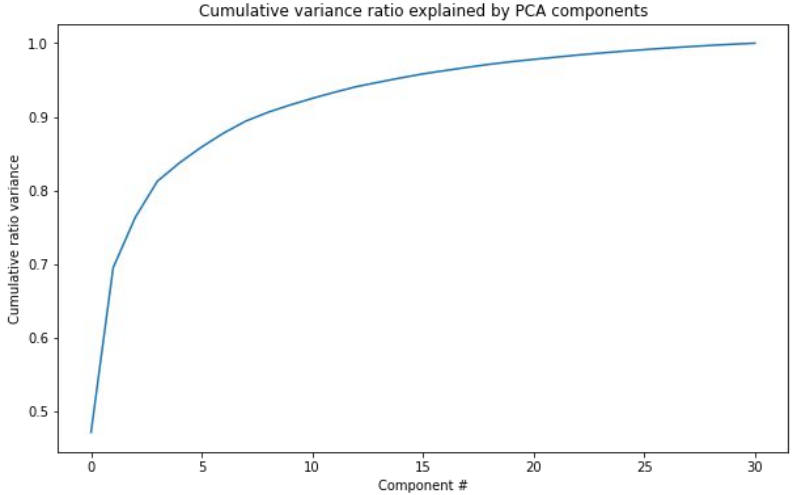
- 788 occupations
- 35 skills
- 379 locations

States


Mean salary per state
Overall mean
Mean Salary
80000
70000
60000
50000
40000
30000
20000
10000
0
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA PR RI SC SD TN TX UT VA VT WA WI WV WY
State

PCA


Explained Variance
Components
0
1
2
3
4
0
5
10
15
20
25
30


Cumulative variance ratio explained by PCA components
Cumulative ratio variance
Component #
0.5
0.6
0.7
0.8
0.9
1.0
0
5
10
15
20
25
30

Baseline model

Uses mean as the best estimate

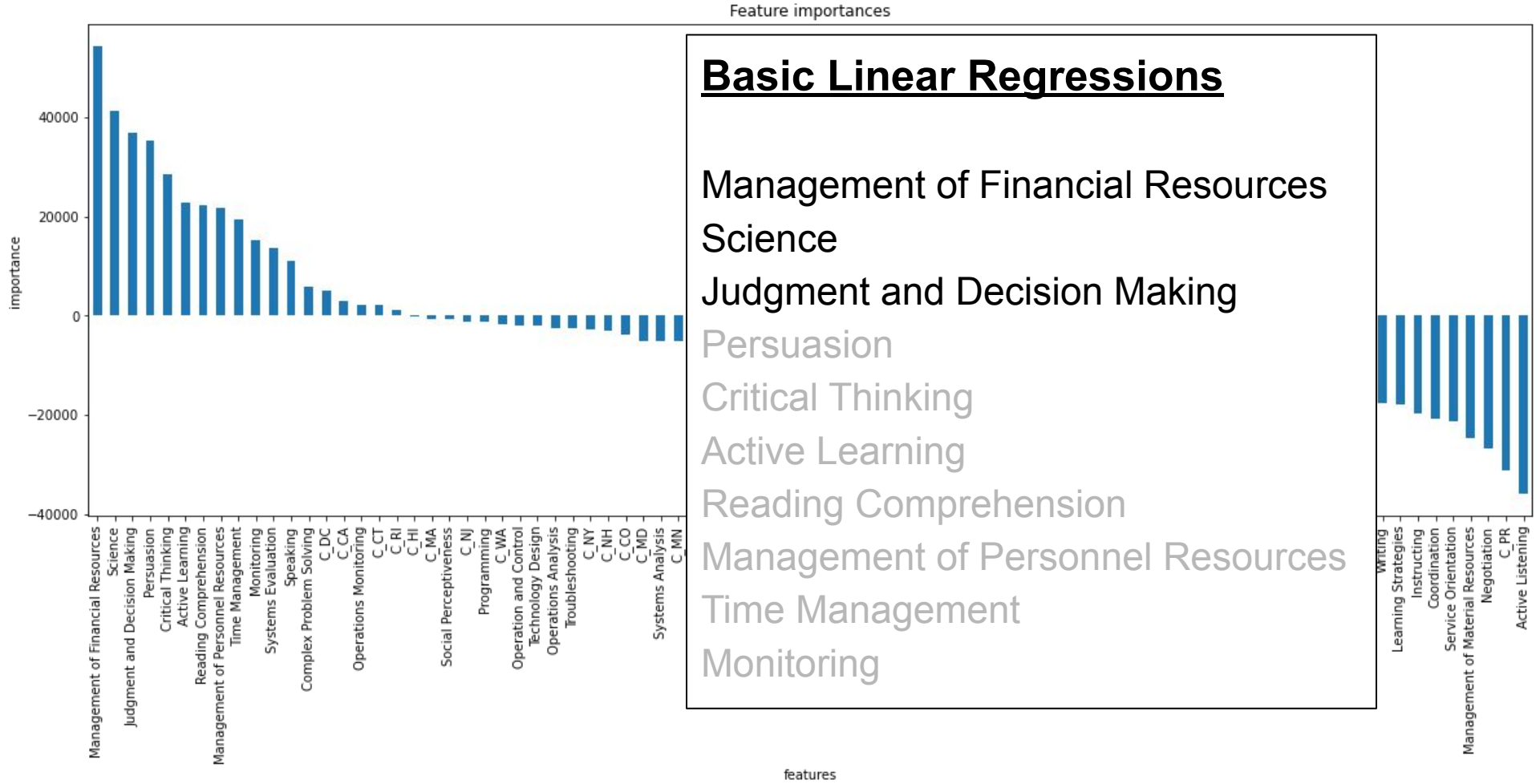
Get basically zero for R^2 error: the model does a terrible job

Fit time = 0.5

Score time = 0.08

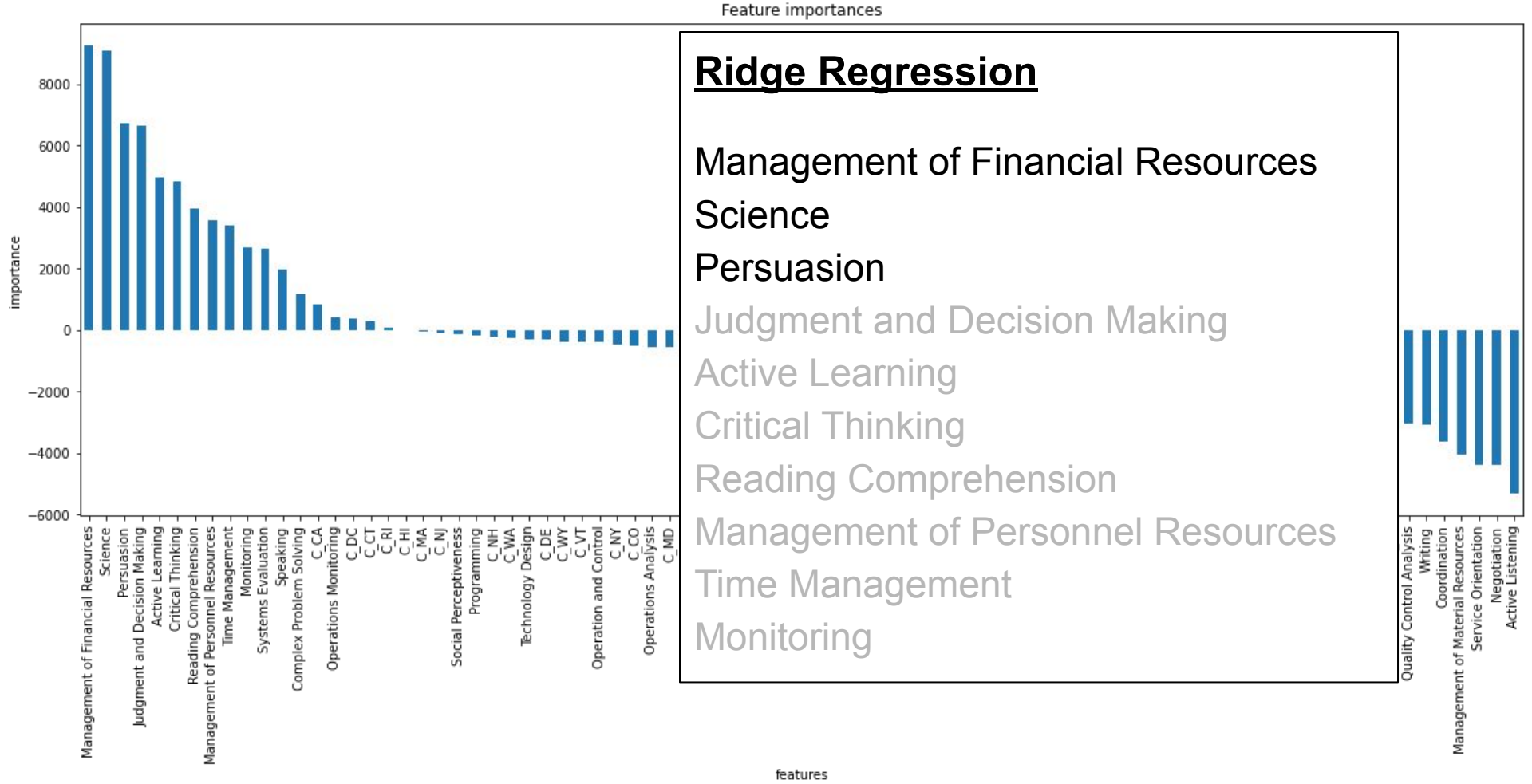
Regression Models

	OLS	Ridge	SVR
R-squared	0.606	0.606	0.586
Fit time	12.88	0.93	91.27
Score time	0.58	0.066	0.0198
MAE test	$14,251		



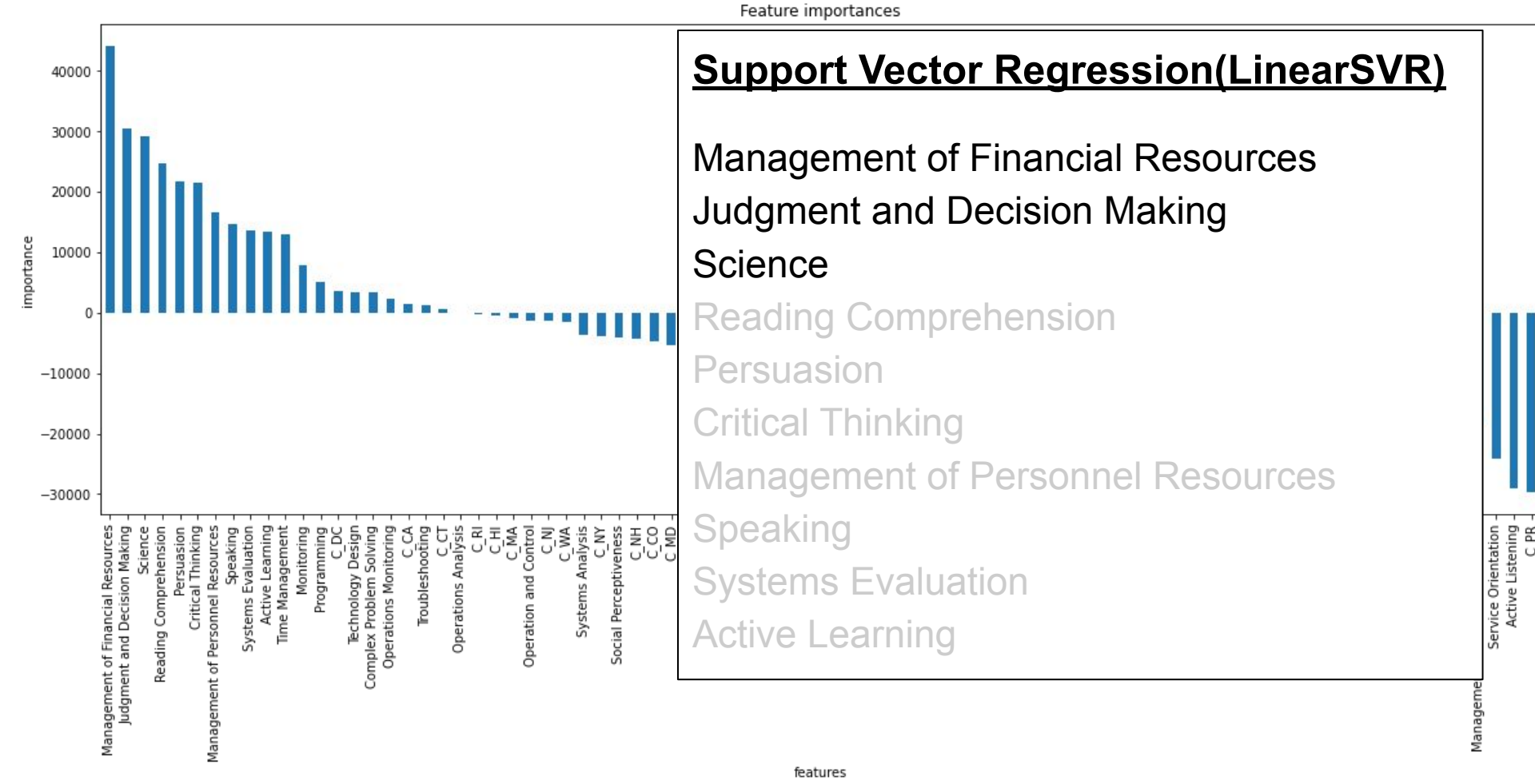
Basic Linear Regressions

Management of Financial Resources
Science
Judgment and Decision Making
Persuasion
Critical Thinking
Active Learning
Reading Comprehension
Management of Personnel Resources
Time Management
Monitoring



Ridge Regression

Management of Financial Resources
Science
Persuasion
Judgment and Decision Making
Active Learning
Critical Thinking
Reading Comprehension
Management of Personnel Resources
Time Management
Monitoring



Support Vector Regression(LinearSVR)

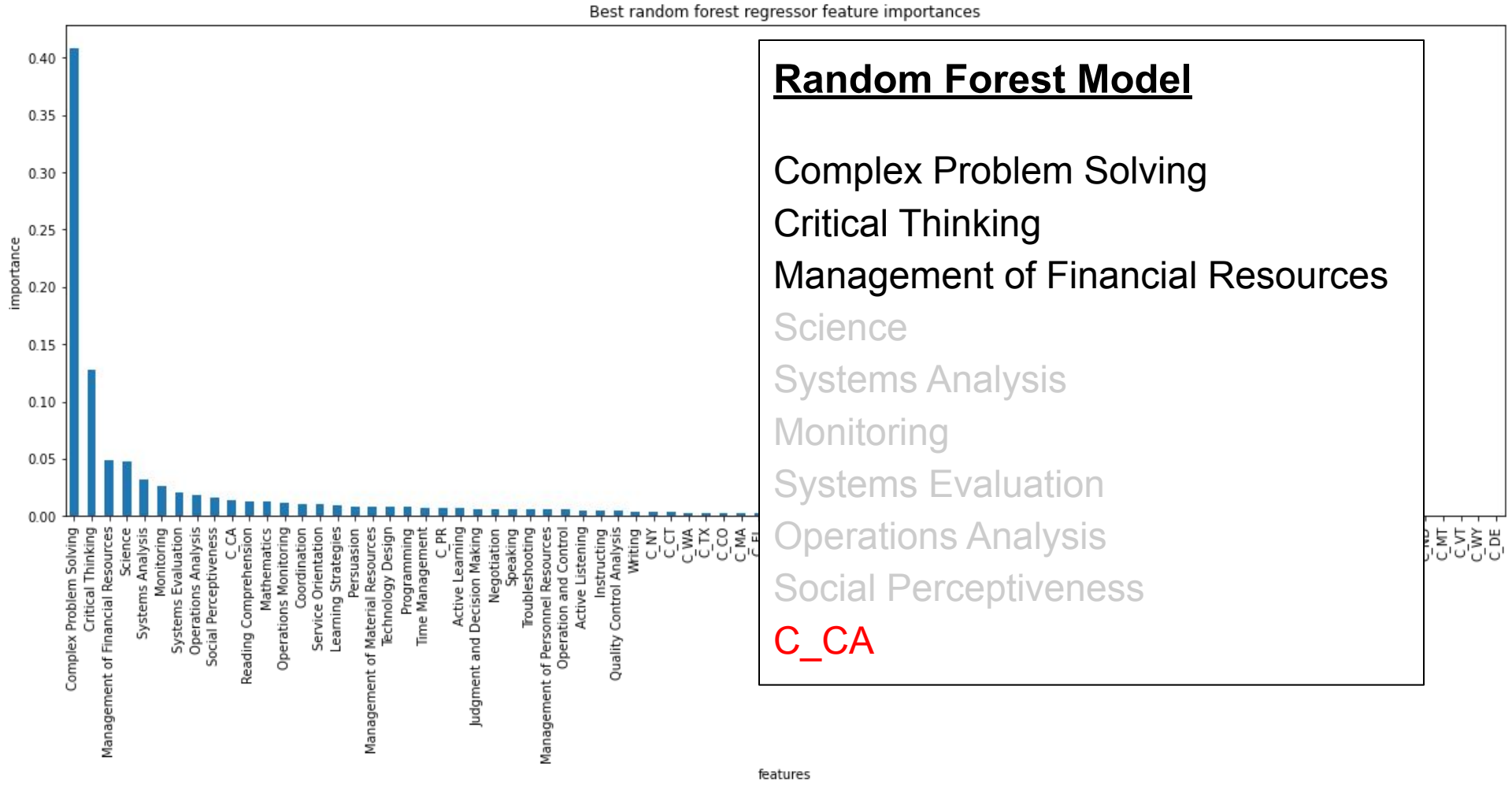
Management of Financial Resources
Judgment and Decision Making
Science
Reading Comprehension
Persuasion
Critical Thinking
Management of Personnel Resources
Speaking
Systems Evaluation
Active Learning

Random Forest Model

(0.8875267860497988, 0.0031824517006334503)

Fit time: 92.8323

Score time: 0.5231



Random Forest Model

Complex Problem Solving

Critical Thinking

Management of Financial Resources

Science

Systems Analysis

Monitoring

Systems Evaluation

Operations Analysis

Social Perceptiveness

C_CA