



Machine Learning for Community Forecasting

Local Indicators Education, Experience, Safety... Locations Southeast, USA Industries Employment Levels Output Results Download Your Campaign

Machine Learning for Community Forecasting

Our app helps organizations design volunteer rewards that to engage volunteers in communities that are the most at risk of decline.

Based on your campaign settings, we'll form clusters of matching locations using local job and income levels, hiring patterns for multiple industries, work experience, education, crime stats, and poverty levels for adults, children and seniors.

View community forecasts... Zip Code GO

1 Design a Campaign

We'll guide you through setting priorities and choosing clusters to create an incentive program that serves communities at risk of decline.

Where would you like to focus your efforts?

- Education Needs
- Work Experience Needs
- Crime / Public Safety Needs
- Youth Below Poverty Level
- Adults Below Poverty Level
- Seniors Below Poverty Level

NEXT

Predictors of Increased Poverty

Random Forest Classifier

features	importance
1 population	0.078512
2 education	0.073835
3 poverty_18to65	0.072386
4 work_experience	0.070808
5 poverty	0.070172
6 working_fulltime	0.069096
7 poverty_under18	0.068506
8 JobsTotal	0.066695
9 poverty_over65	0.059081
10 JobsConstruction	0.048336
11 JobsHealthcare	0.042636
12 JobsManufacturing	0.041522
13 JobsProfessional	0.041481
14 JobsTransport	0.040196
15 JobsTrade	0.038642
16 JobsRealestate	0.034896
17 JobsEntertainment	0.0324
18 working_fulltime_poverty	0.030949
19 JobsAgriculture	0.019851

2016 predicting 2017
Feature importance in zipcodes where poverty increased by 2% the next year.

```

graph TD
    TrainingSet[Training Set] --> TrainingSample1[Training Sample 1]
    TrainingSet --> TrainingSample2[Training Sample 2]
    TrainingSet --> TrainingSampleN[Training Sample n]
    TrainingSample1 --> DecisionTree1[Decision Tree 1]
    TrainingSample2 --> DecisionTree2[Decision Tree 2]
    TrainingSampleN --> DecisionTreeN[Decision Tree n]
    DecisionTree1 --> Voting[Voting]
    DecisionTree2 --> Voting
    DecisionTreeN --> Voting
    Voting --> Prediction[Prediction]
    TestSet[Test Set]
  
```

Best Params: max depth: 8; n-estimators: 100
Accuracy before tuning: 69%. Accuracy after tuning: 71%.

Local Indicators Education, Experience, Safety... Locations Southeast, USA Industries Employment Levels Output Results Download Your Campaign

Education Needs

0 25 50 75 100

Work Experience Needs

0 25 50 75 100

Youth Poverty

0 25 50 75 100

Adult Poverty

0 25 50 75 100

Senior Poverty

0 25 50 75 100

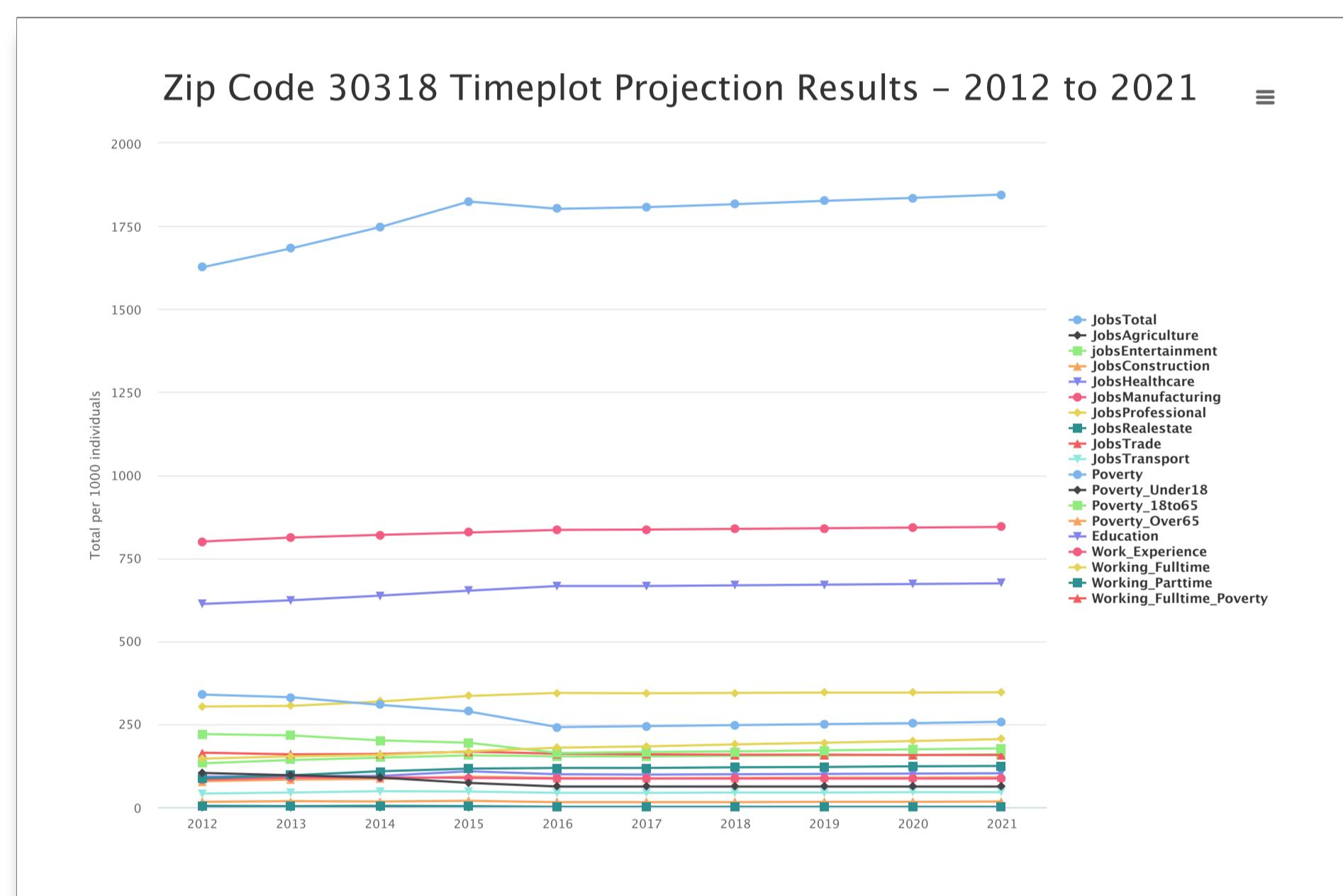
Zip code map of Georgia showing poverty clusters. Cluster 1 (orange) covers the northern part of the state, while Cluster 2 (blue) covers the southern part.

Cluster 1

- num: 626
- Poverty_Under18_stats:
 - min: 0
 - max: 168
 - mean: 60.966453674121404
 - variance: 1430.0100567526456
- Poverty_Over65_stats:
 - min: 0
 - max: 39
 - mean: 15.087859424920127
 - variance: 99.38045963519068

Cluster 2

- num: 81
- Poverty_Under18_stats:
 - min: 0
 - max: 430
 - mean: 100.4074074074074
 - variance: 6583.6982167352535
- Poverty_Over65_stats:
 - min: 0
 - max: 184
 - mean: 45.65432098765432
 - variance: 1298.9422344154852



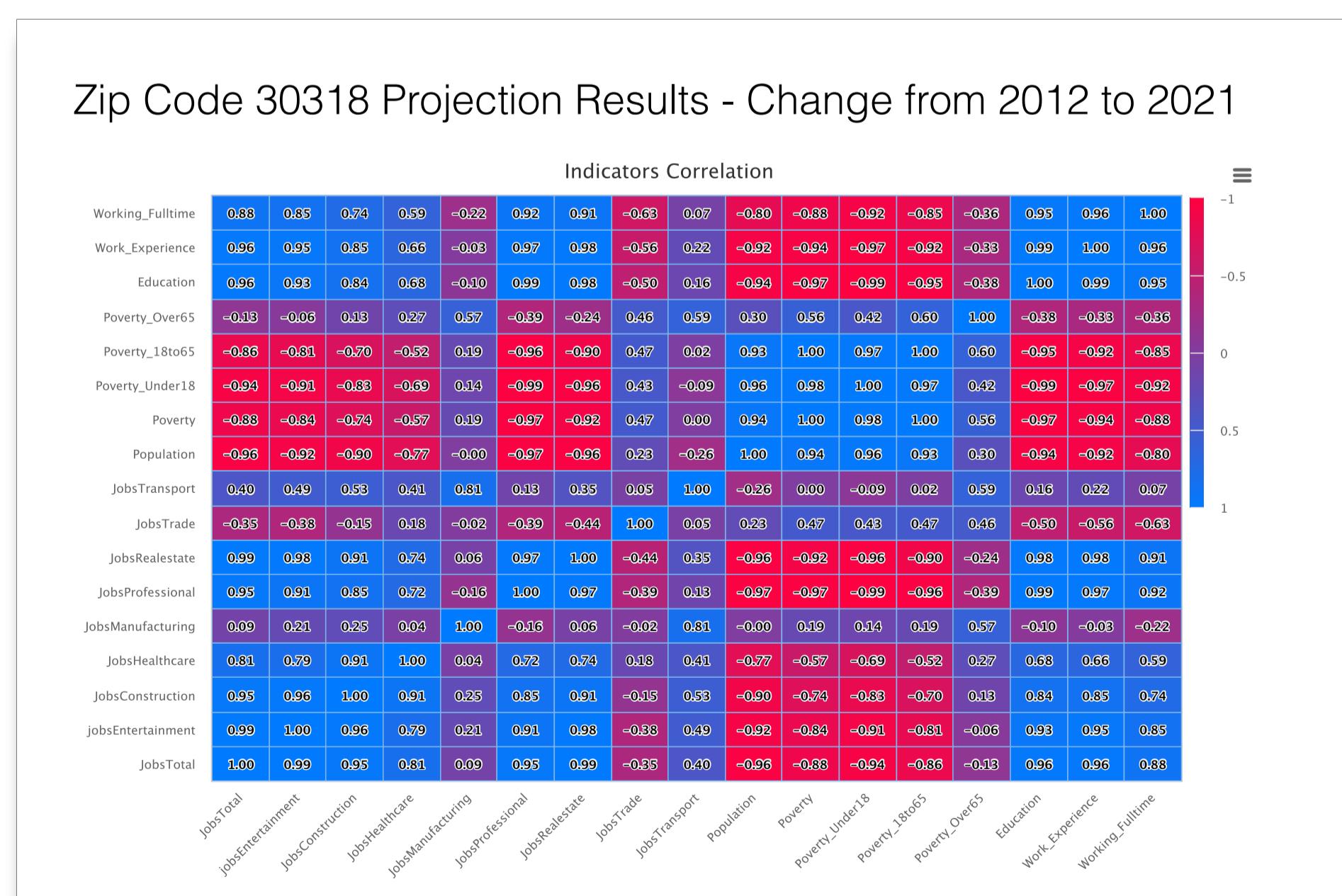
Local Indicators Education, Experience, Safety... Locations Southeast, USA Industries Employment Levels Output Results Download Your Campaign

Machine Learning for Community Forecasting

Our community forecasting model is trained using time series Census and FBI data from 2012 to 2016. Each indicator's performance is compared to prior years to generate projected estimates for the next 5 years. Projections are based on a cross validated **Lasso Regression** via the Least Angle Regression (LARS), forward-selection algorithm which helps automatically optimize both feature weightings and the regularization parameter. We use this approach to improve generalizability and prevent potential overfitting caused both by local anomalies and short-term historical trends.

View community forecasts... Zip Code GO

30318



Team 6 - datascape.github.io/community

Loren Heyns, Huyen Nguyen, Zhengfu Li, Truc Pham, Christopher Fleisher, Trai Tran
Computational Science & Engineering Instructor Polo Duen Horng Chau
Georgia Tech CSE 6242: Data and Visual Analytics (DVA) - Spring 2019

Fun facts about our forecasting tools

We use the highly efficient **DBSCAN** clustering algorithm to generate map clusters in realtime. This poster is output from a giant webpage designed to create **custom posters for zipcodes**. We developed a **JSON Hierarchy Reader** that converts lists to HTML for our map cluster buckets. We use the same reader with the **DiffBot Knowledge Graph API** to list industries by zip code. Our GitHub pages use **Markdown-to-HTML** to build webpages from multiple files. Our tables are displayed using **CSV-to-HTML** via **D3**. Our maps use **LeafletJS** with **D3 layers**.