Wages Prediction: Project 4 Hackathon
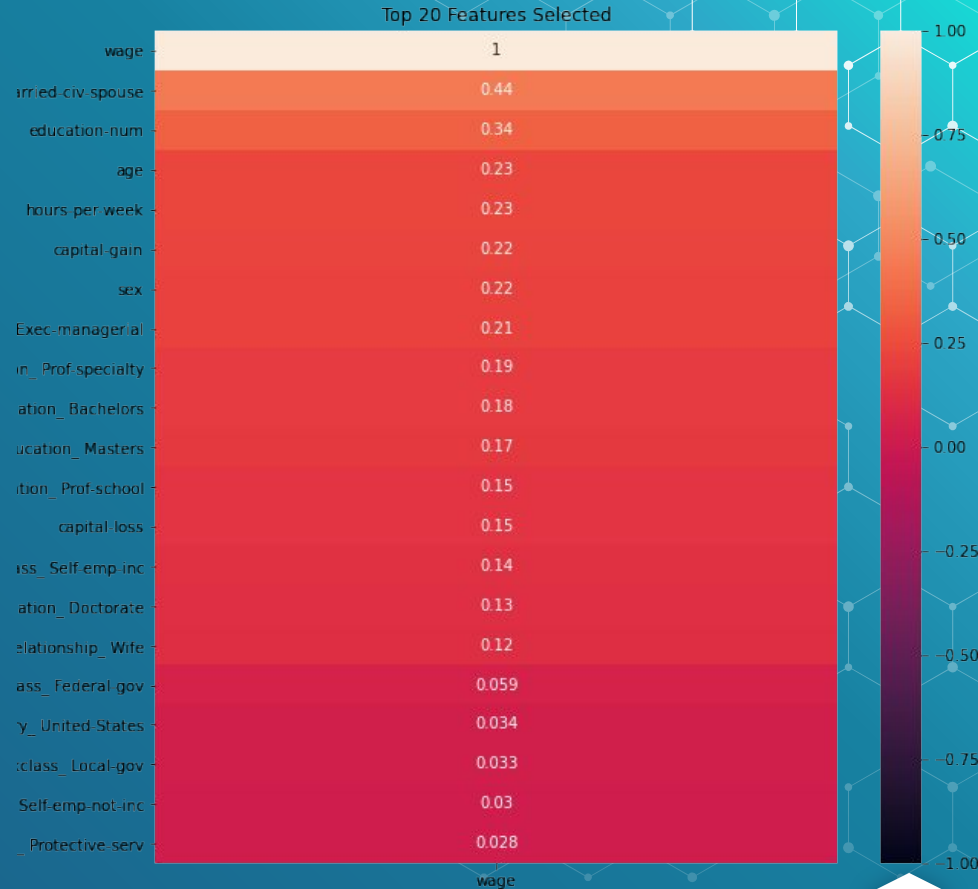
PROJECT INTRO

Katharine King

N.A. Kovacs

Will Hanley

Young Park

DATA CLEANING

- Limitations: 20 features

- Dummify & order by correlation

- Keep top 20 highest correlations

Top 20 Features Selected

Feature	wage
wage	1
married-civ-spouse	0.44
education-num	0.34
age	0.23
hours-per-week	0.23
capital-gain	0.22
sex	0.22
Exec-managerial	0.21
n_ Prof-specialty	0.19
ation_ Bachelors	0.18
ucation_ Masters	0.17
ation_ Prof-school	0.15
capital-loss	0.15
ss_ Self-emp-inc	0.14
ation_ Doctorate	0.13
elationship_ Wife	0.12
ass_ Federal-gov	0.059
y_ United-States	0.034
class_ Local-gov	0.033
Self-emp-not-inc	0.03
_ Protective-serv	0.028

MODELS

Numerous models gridsearched for the best outcomes.

Initial Exploration

Getting a model set up:

Logistic Regression (81%)

Additional explorations kept LogReg at the top of the leaderboard for some time (85%)

Exploring Polynomial Features

ADDITIONAL MODELS

	Training	Testing
Random Forest	0.526	0.488
KNN	0.504	0.374
SVM	0.859	0.854

FINAL SUCCESS

Adaboost:
- Test: 0.862
- Train: 0.869

XGBClassifier:
- Test: 0.879
- Train: 0.872

THANKS!

Open for Questions & Comments
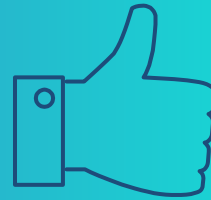
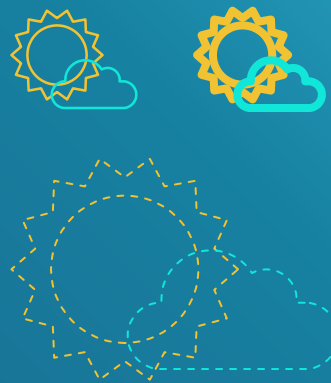*Katherine King, N.A. Kovacs, Will Hanley, Young Park*

8

SlidesCarnival icons are editable shapes.

This means that you can:

- ◇ Resize them without losing quality.
- ◇ Change fill color and opacity.
- ◇ Change line color, width and style.

Isn't that nice? :)

Examples:

Find more icons at
slidescarnival.com/extra-free-resources-icons-and-maps

9