Winning Characteristics in Technology Competitions

Rita Savill
EPPS 6323

Research Purpose

To explore characteristics of Hackathon projects and try to predict the probability of winning.

Inspired by sports analytics.

Hackathons are competitive coding events that generally take place over a 24-48 hour period.

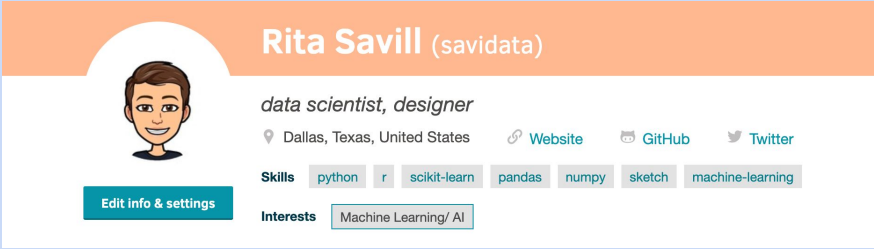
Characteristics of the projects and the people who made them (‘player stats’)

Data Collection

Projects from MLH sponsored hackathons in North America for the 2019 season for which the projects are viewable on devpost.com

Scraped from devpost.com using Python's BeautifulSoup library

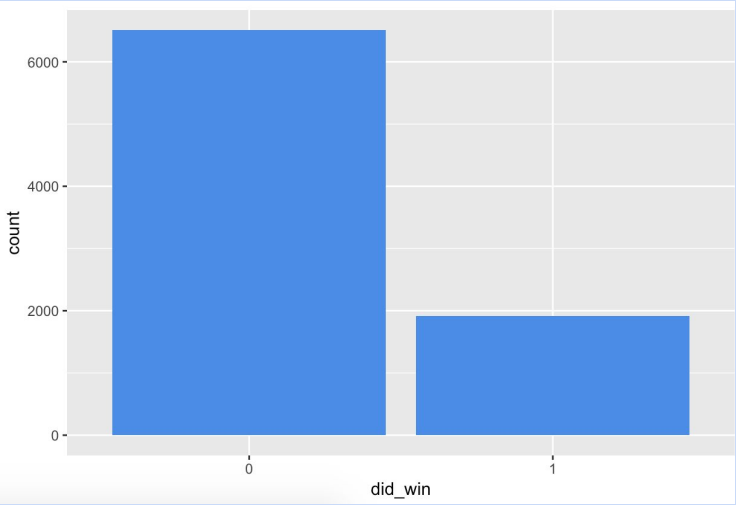
8,422 observations; 1 binary outcome variable, 13 predictor variables

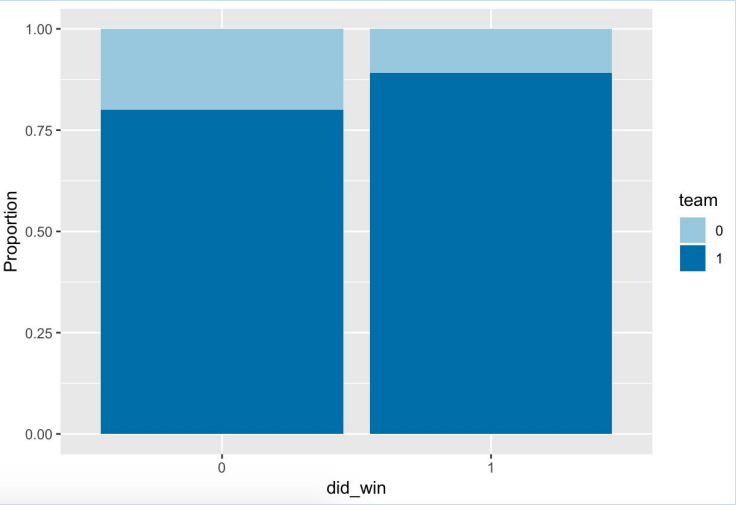


EDA

Missing values: 585 instances from the text columns

Data is imbalanced





Text Mining

Frequency of built-with-tags

```
# A tibble: 795 x 2
   tags            n
   <chr>       <int>
 1 javascript   2747
 2 python       2629
 3 html         1597
 4 css          1530
 5 node.js      1233
 6 google-cloud 1058
 7 java          969
 8 html5         917
 9 firebase      802
10 flask         794
# … with 785 more rows
```

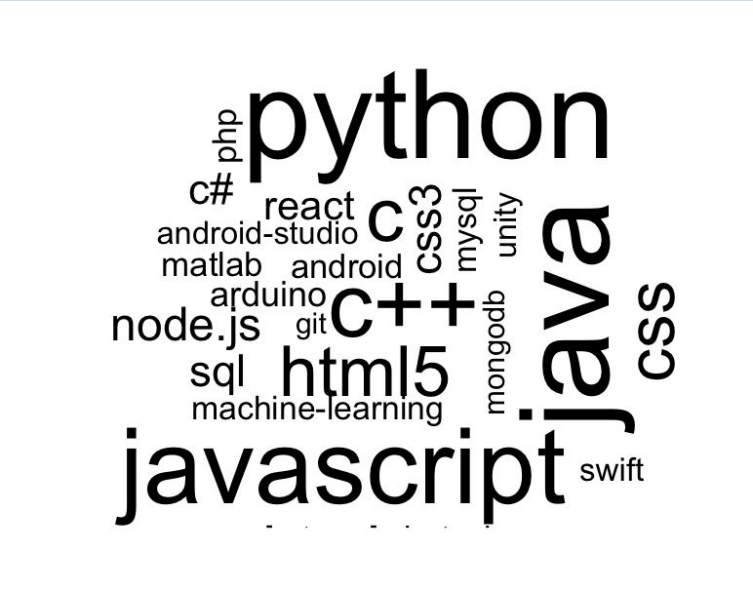
Frequency of unique skills per team

```
# A tibble: 4,496 x 2
   skills        n
   <chr>     <int>
 1 java       5485
 2 python     5469
 3 javascript 4517
 4 c++        3793
 5 c          3038
 6 html5      2752
 7 css        2619
 8 html       2383
 9 node.js    1573
10 css3       1482
# … with 4,486 more rows
```

TAGS


javascript
express.js
html5
node.js
arduino
c#
azure
unity
flask
css3
android-studio
android
twilio
react
swift
google-cloud
html
mongodb
firebase
bootstrap
java
google-maps
python

SKILLS


python
php
c#
react
c
android-studio
matlab
android
arduino
node.js
git
c++
sql
html5
machine-learning
mongodb
mysql
unity
java
javascript
swift

Classification Models

Logistic Regression

- Accuracy(balanced): 0.8744
- Specificity: 0.9608
- Sensitivity: 0.7880

```
          Reference
Prediction    0    1
         0 1251   81
         1   51  301
```

Random Forest

- Accuracy(balanced): 0.8785
- Specificity: 0.9639
- Sensitivity: 0.7932

```
          Reference
Prediction    0    1
         0 1255   79
         1   47  303
```

Future Research

Expansion on this analysis

Explore gender ratio

Relationships between text tags/skills

Time series

https://www.instagram.com/p/BxNwf9fF1sA/