





```
$ sudo apt-get install libfontconfig
$ sudo apt-get install libatlas-base-dev gfortran
$ sudo apt-get install nodejs
$ sudo apt-get install npm
$ sudo apt-get install nodejs-legacy
$ sudo npm -g install phantomjs
$ sudo apt-get install python-pip
$ sudo apt-get install python-dev
$ sudo pip install numpy
$ sudo pip install pandas
$ sudo pip install scikit-learn
$ sudo pip install scipy
...

$ python scrape-robot.py < China-2015.txt > China-2015-log.txt
...

Traceback (most recent call last):
  File "scrape-robot.py", line 310, in <module>
    scrape-instance.xml-scrape()
  File "scrape-robot.py", line 50, in xml-scrape
    self.parse-text() File "scrape-robot.py", line 175, in parse-text
    while (xml-source.find('<?xml') != 0) and ct <= 5:
AttributeError: 'NoneType' object has no attribute 'find'
```



Ivan Echevarria

✓ Scrape Approved

Personal Menu

🎓 Profile

🕒 History

🚪 Sign Out

Main Menu

💎 Process Data

🔗 Classifier Manager

📘 Getting Started <

📖 Guac Wiki

💬 Contact the Team

1. Data Source

☒ Scrape From Factiva☒ Build TUFF Query☐ Manual Query☒ Scrape Now☐ Schedule Scrape☐ Upload Articles

2. Data Processing

☒ Classify

TUFF ▾

Yes ▾

🗑 Remove

Conflict ▾

Yes ▾

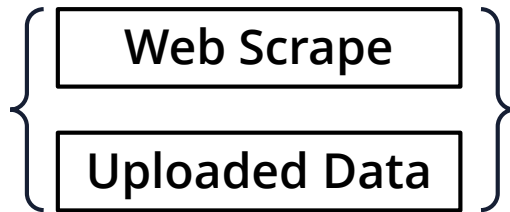
🗑 Remove

+ Add Classifier

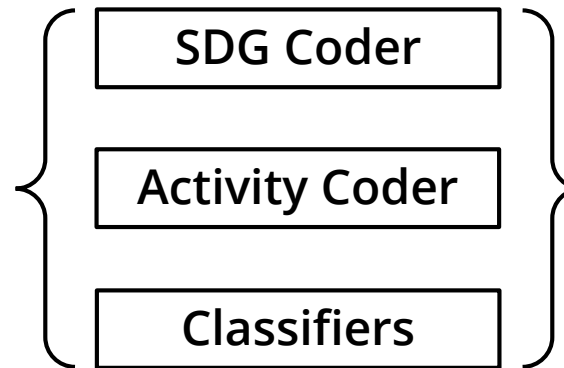
☒ Activity CodeYou will receive a **.csv** file with the activity codes associated with each of your documents



1. Get Data

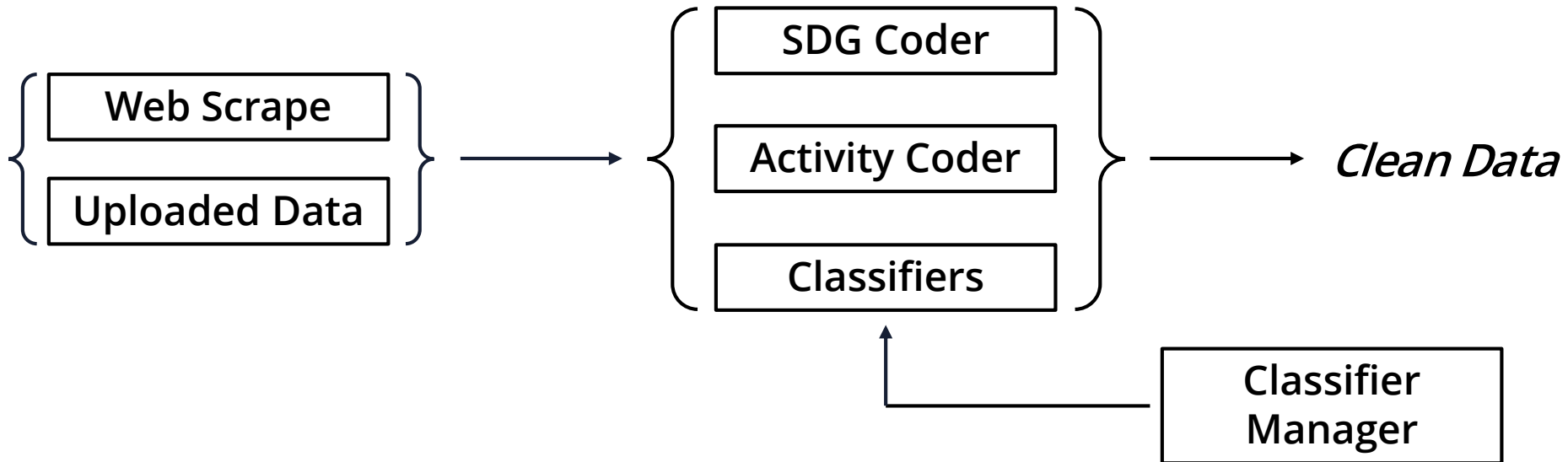


2. Process Data



Clean Data

Classifier
Manager





Dan Jelf

full stack developer;
maintains autocoders



Nadia Aly

machine learning wizard;
builds classifiers



Ivan Echevarria

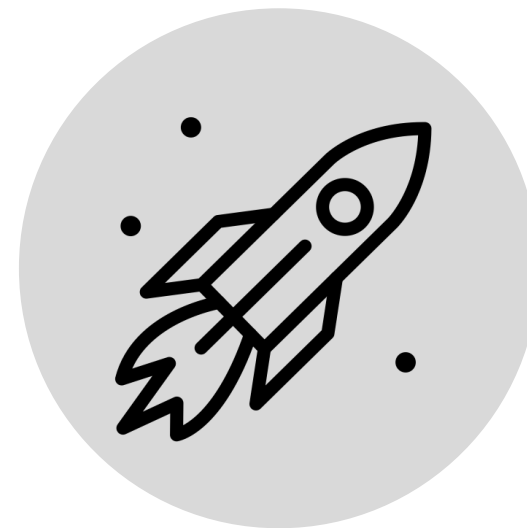
frontend developer;
operates scraper



Proven Tech

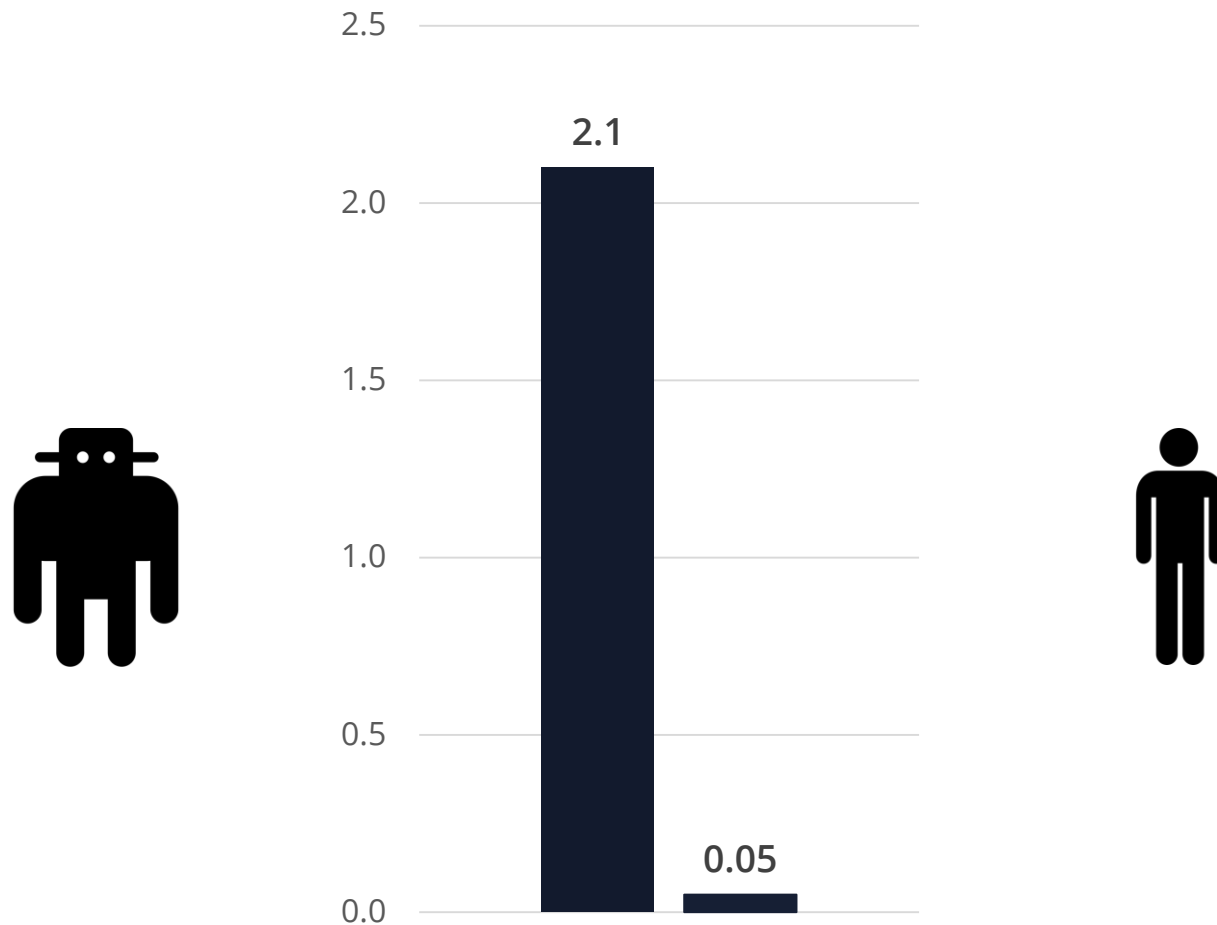


Proven People

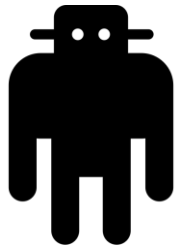


High Potential

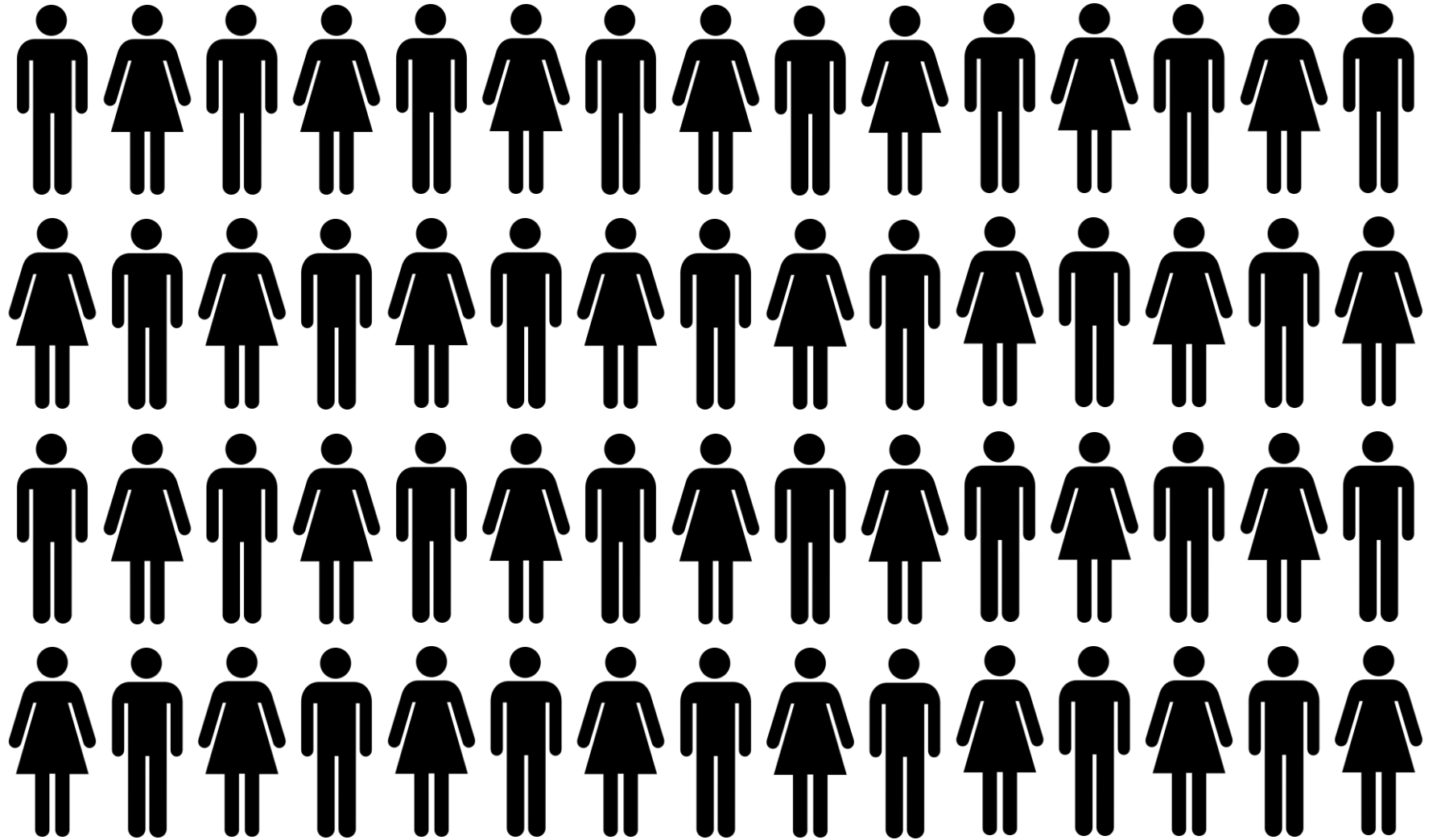
Volume: 2.1 million+



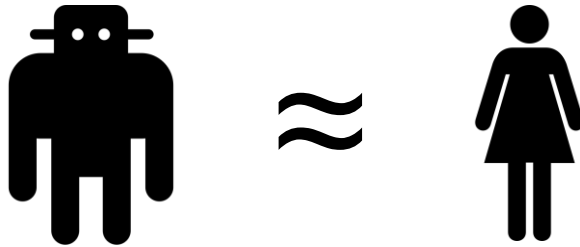
Velocity: 250+ each minute



\geq



Veracity: 98%+





Presentation Credits

AlGA Man Icon by **Edward Boatman**

AlGA Woman Icon by **Edward Boatman**

Robot Icon by **Creative Stall**

Gears Icon by **Márcio Duarte**

Weight Lift Icon by **Andrew Doane**

Rocket Ship Icon by **AB**

Open Sans Font by **Google Corporation**

All other materials © Guac 2016