Subreddit Modeling

Michael Mejia

Data Gathering

- Pushshift.io
  - **Jason Baumgartner**

Artificial and MachineLearning

- **Baseline Score**
  - **0 0.509648**
  - **1 0.490352**
- **Quantity of Data**
  - **10678**

CountVectorizer Vs **TFIDFVectorizer**

- Naive Bayes
  - Multinomial
  - Gaussian
  - Bernoulli

Decision Tree

Random Trees?

Best Model so far

Logistic regression