

# Reddit Classification

## Cycling Posts

Russell Garner

# Questions

- **Can a model classify mountain biking and road cycling post?**
- **Is it better than guessing the majority class?**

# Traditional



# Tech



# **Cycling Industry Estimated Value 6.2 Billion Dollars**

# Sub Reddits

- **Velo**

<https://www.reddit.com/search/?q=velo>

your place for competitive amateur racing discussion & information!

- **MTB**

<https://www.reddit.com/r/MTB/>

A cycling community enthusiastic about mountain biking and all things related.

# Models

## Logistic Regresson

- **Train Accuracy Score 0.893**
- **Test Accuracy Score 0.887**

## Multinomial Naive Bayes

- **Train Accuracy Score 0.888**
- **Test Accuracy Score 0.880**

## Random Forest

- **Train Accuracy Score 0.876**
- **Test Accuracy Score 0.858**

# Null vs Model

## Null

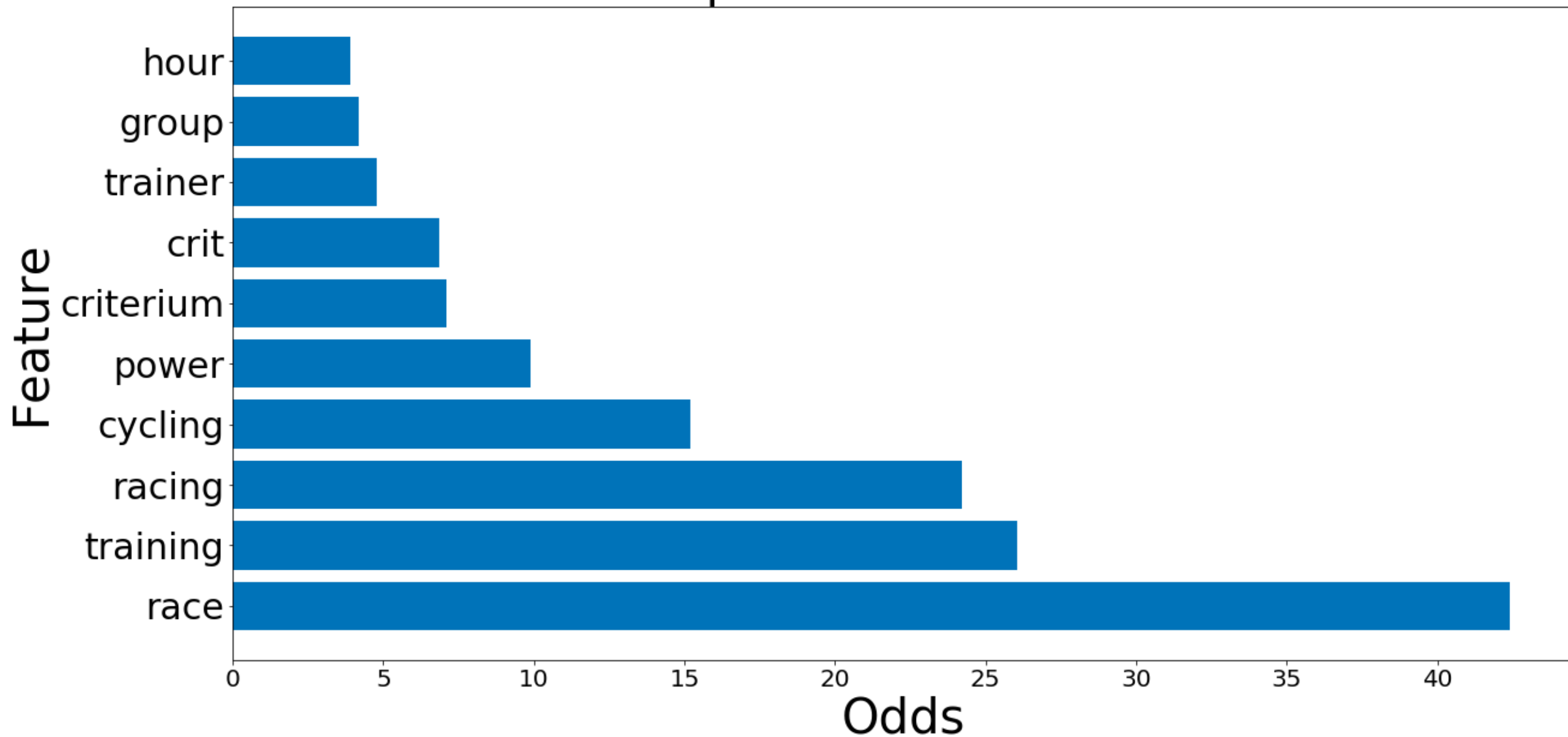
- **Accuracy Score 0.519**

## Logistic Regression

- **Test Accuracy Score 0.887**

# Top Words for Road

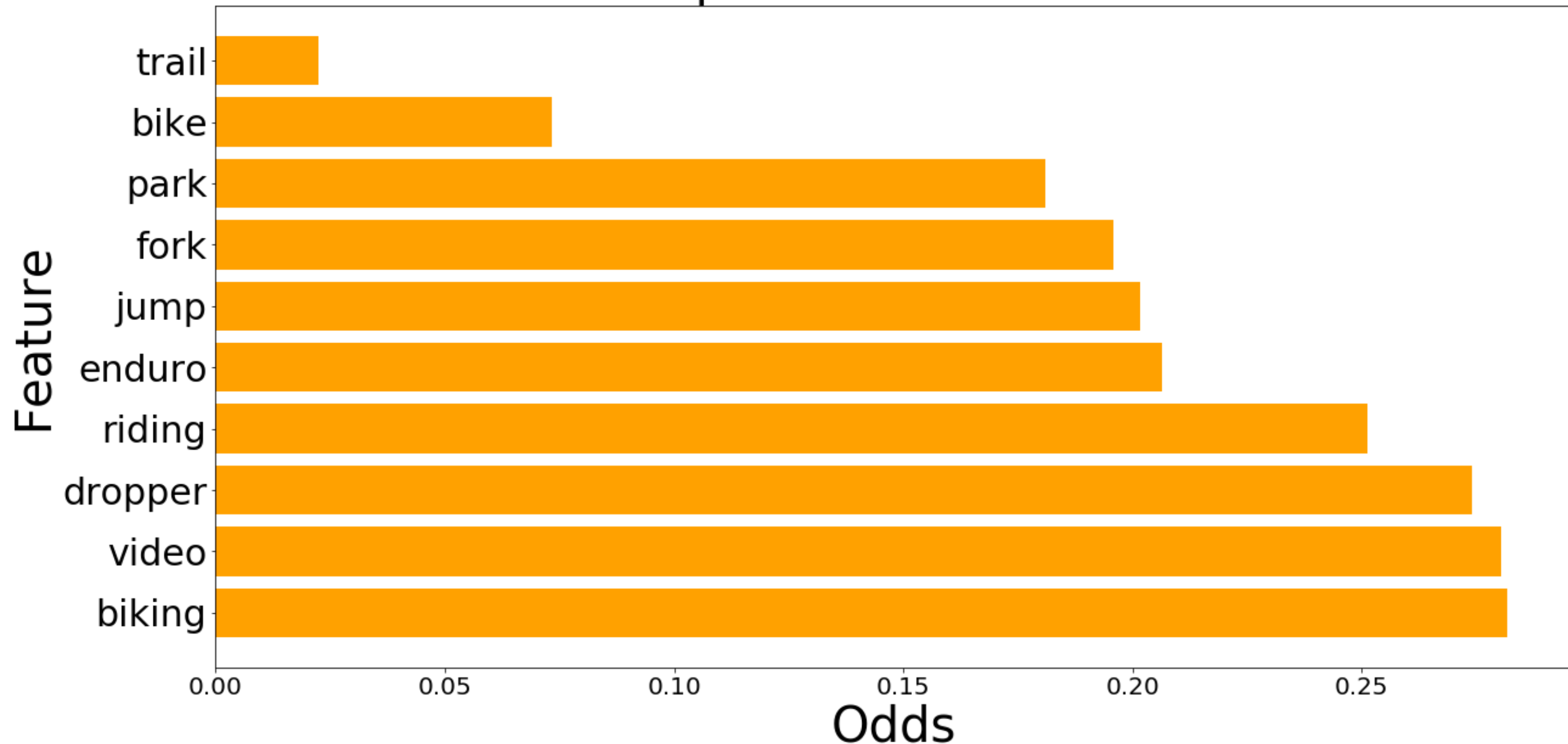
Top 10 Velo Features





# Top Words for Mountain

Top 10 MTB Features



# Classification

PRED MTB

PRED VELO

ACTUAL MTB

190

27

ACTUAL VELO

24

210

# Metrics

Sensitivity	0.90
-------------	------

Specificity	0.88
-------------	------

Precision	0.88
-----------	------

Misclass	0.11
----------	------



# Questions



[https://en.wikipedia.org/wiki/Sam\\_Hill\\_\(cyclist\)](https://en.wikipedia.org/wiki/Sam_Hill_(cyclist))

<https://www.cyclingweekly.com/news/racing/peter-sagan-easy-reach-certain-level-much-difficult-stay-410540>