

# David Arnold

Data Scientist and Engineer

## MACHINE LEARNING TO PREDICT A POPULAR MEME ON REDDIT



```

    'width': 140},
    'created_utc': 1517845675.0,
    'author': 'arg6531',
    'id': '7vflh6',
    'ups': 38049,
    'downs': 0,
    'media': 'https://i.redd.it/4vndclxu0fe01.jpg'},
    '99': {'title': 'SP [REDACTED] KY',
    'thumbnail': {'thumbnail': 'https://a.thumbs.redditmedia.com/i08HyxyYkFKNNOPhqmS-uTcQj-rsBlzwjpuTZ0T4yZ0.jpg',
    'height': 139,
    'width': 140},
    'created_utc': 1509430037.0,
    'author': '_Vice_roy_',
    'id': '79tx0i',
    'ups': 37795,
    'downs': 0,
    'media': 'https://i.redd.it/xqzmabltw3vz.jpg'},
    '100': {'title': 'What the fuck am i supposed to put here?',
    'thumbnail': {'thumbnail': 'https://b.thumbs.redditmedia.com/V14e6-numhlMHY0CVM-E2PVtZ7rXDK2iq6v39LFW_4M.jpg',
    'height': 140,
    'width': 140}
  }
}

```



	title	created_utc	author	id	ups	downs	media	thumbnail.thumbnail	thumbnail.height	thumbnail.width
0	Number One	1.502621e+09	DrarenThiralas	6tehbc	87082	0	https://i.redd.it/7wgs4dkiihfz.png	https://b.thumbs.redditmedia.com/FAS_fwvpmzuP...	121	140
1	Got'em	1.523557e+09	CasualDad8675309	8bse8k	75251	0	https://i.redd.it/65bzzi0isir01.jpg	https://b.thumbs.redditmedia.com/hulqG6r8L3pry...	98	140
2	50-0	1.503862e+09	NikiTosThePleb	6we7gp	64236	0	https://i.redd.it/19c4ggoz0ciz.png	https://b.thumbs.redditmedia.com/1LvJdhkQtbtvZ...	140	140
3	Allow	1.515929e+09	bananahands0666	7qbcqw	63386	0	https://i.redd.it/qt5p8ozio0a01.png	https://a.thumbs.redditmedia.com/OoC51rX7qUafr...	110	140
4	*mild concern*	1.516919e+09	DJ_MilkBags	7szw80	61458	0	https://i.redd.it/1cudu2jlgac01.jpg	https://b.thumbs.redditmedia.com/f1HuHxnds0Lgb...	134	140

# Features / Target

## ⦿ Features

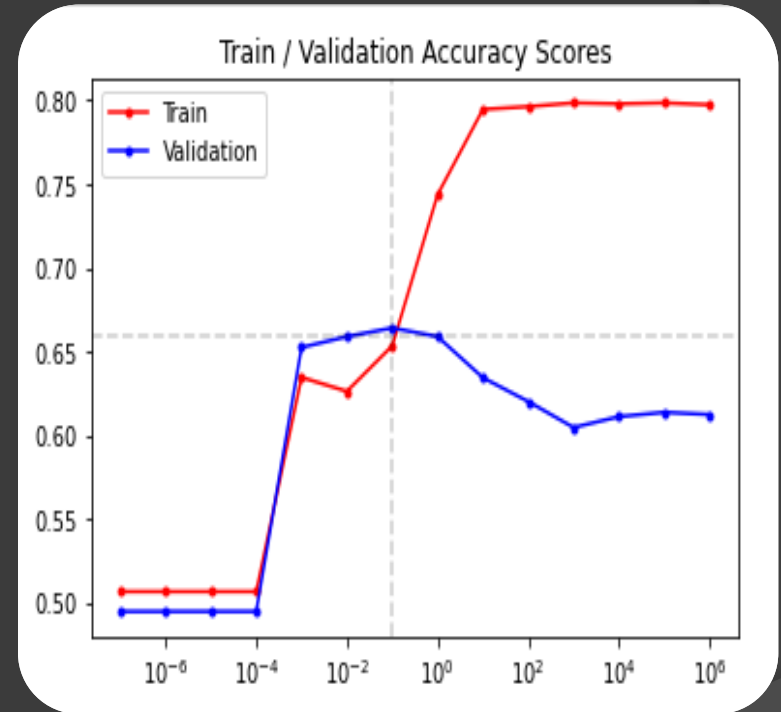
- Date – Years, Months, Days, Weekends
- Title – Vectorized key words
- Author – Binned by post amount
- Thumbnail Image Area

## ⦿ Target

- Up Votes
- Binomial Classification
  - “Good” More than 24000 votes
  - “Bad” Less than 24000 votes

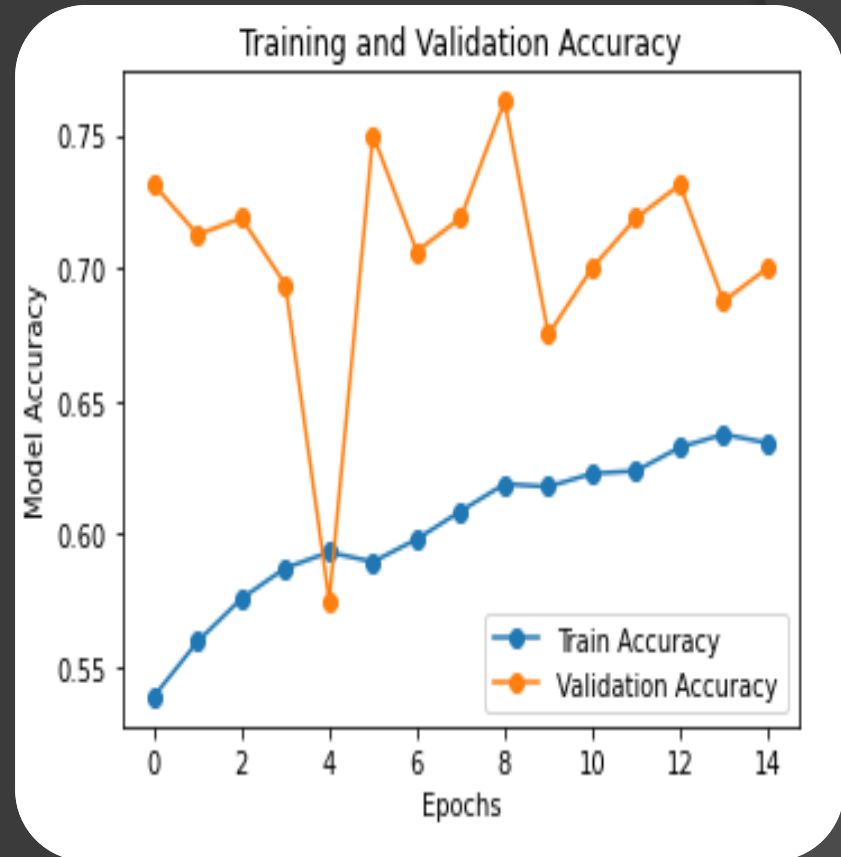
# Model Accuracy – Test Data

- ⦿ Logistic Regression
  - 66.4%
- ⦿ Decision Trees
  - 65.2%
- ⦿ K-Nearest Neighbors
  - 59.1%



# ML Modeling – Convolutional Neural Network (CNN)

- ⦿ Classification
  - Good / Bad
- ⦿ Pre-Trained Model
  - VGG16
- ⦿ Validation Accuracy
  - ~71%



TensorFlow



Keras

# Good Meme Example

**Girls when  
they see a  
spider**



**The spider**



# Good Meme Example

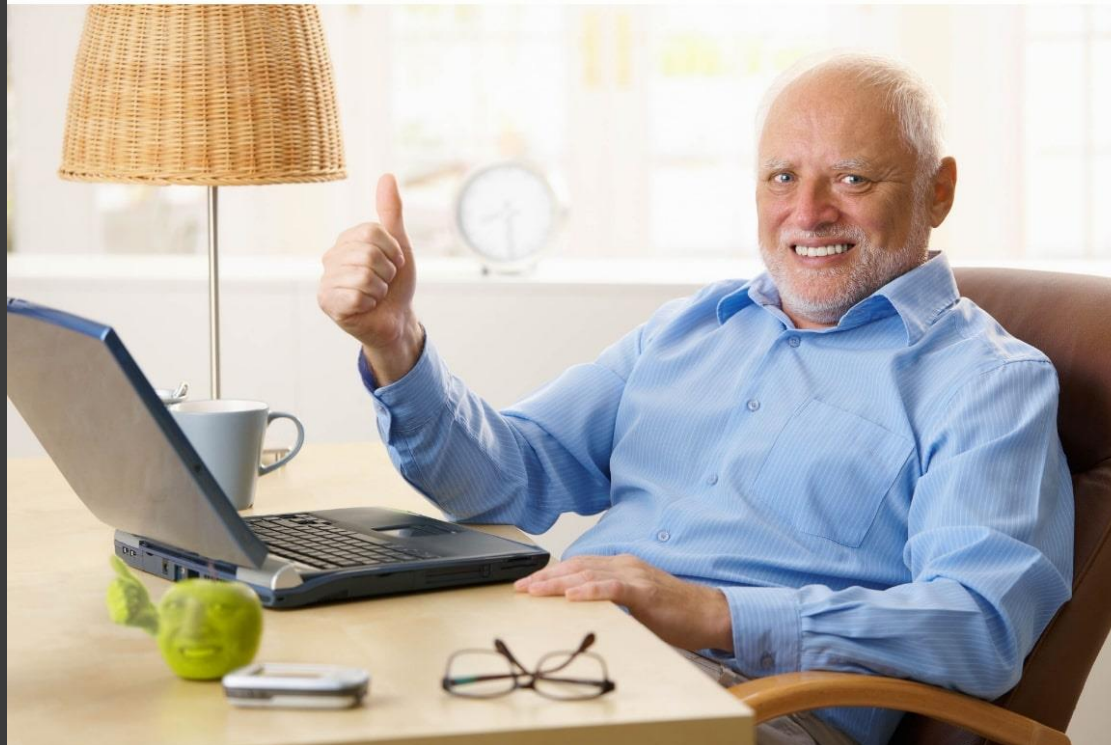
2020 showing 2021 around the workplace





# Bad Meme Example

When your anti-CNN meme only gets 5 upvotes but at least you got 5 people to participate in the Meme War so you're still happy





# Bad Meme Example



# Summary

- ◉ Memes On Social Media
  - Major form of communication
  - Used by Corporations & Social Influencers
  - Societal Trends & Events
- ◉ Test on new Dataset
- ◉ Share on Kaggle



# Ways to Connect:



/dave-m-arnold



orion99da@gmail.com