



Tweet Classification

Using ML and NLP

Predicting tweet sentiment based on content

Table of contents

01

Business Problem +
Project Overview

02 Data Prep / Processing

03

Modeling / Evaluation

04 Conclusion

Business Problem

“Is it possible to utilize **machine learning** and **natural language processing** to predict the **sentiment** of a tweet based on its **content**?”

Project Overview



Data

9000+ classified tweets in their base form



Classification

Positive, Negative, Neutral Sentiment



Models

LogReg, Naive Bayes, Random Forests

Predictability?

Data Prep / Processing

Preprocess

- Converted text to lowercase
- Removed punctuation
 - Stripped words
 - Removed Stopwords



Lemmatize

- Used WordNet Lemmatizer
 - Mapped position tags
 - Tokenized Words



Vectorize

- Used TFIDF Vectorizer



Models + Evaluation

Model Type	Metrics (in %'s)					
	Training Accuracy	Testing Accuracy	Training Recall	Testing Recall	Training Precision	Testing Precision
Reports						
Dummy	60	60	60	60	36	36
Base LogReg	91	67	91	67	91	66
Naive Bayes	75	65	75	65	80	65
Random Forests	77	64	77	64	80	62
LogReg (GridSearch)	65	89	65	89	65	90

Conclusion

Base

Decent model but very overfit

GridSearch Log Reg

Decent model and not overfit!

Random Forests

May prove useful with further tuning [GridSearch]

91/67%

Recall

65/89%

Recall

77/64%

Recall

Thanks!

Do you have any questions?



Mason Walter
Email: masonchasewalter@gmail.com
Github: [@Wingaero](https://github.com/Wingaero)
LinkedIn: [/in/mason-c-walter](https://www.linkedin.com/in/mason-c-walter)

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon and infographics & images by Freepik