



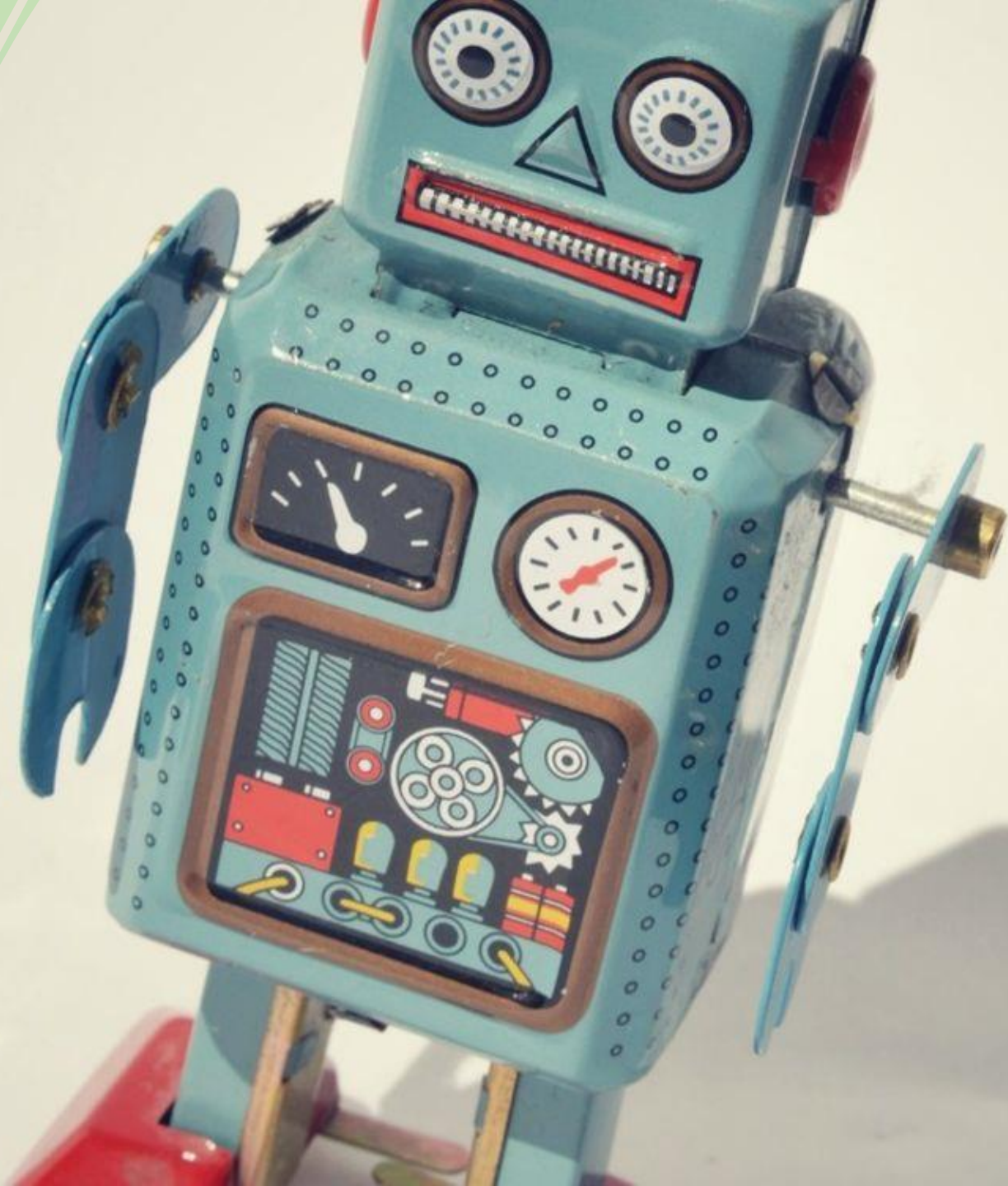
# Business Gains with Natural Language Processing

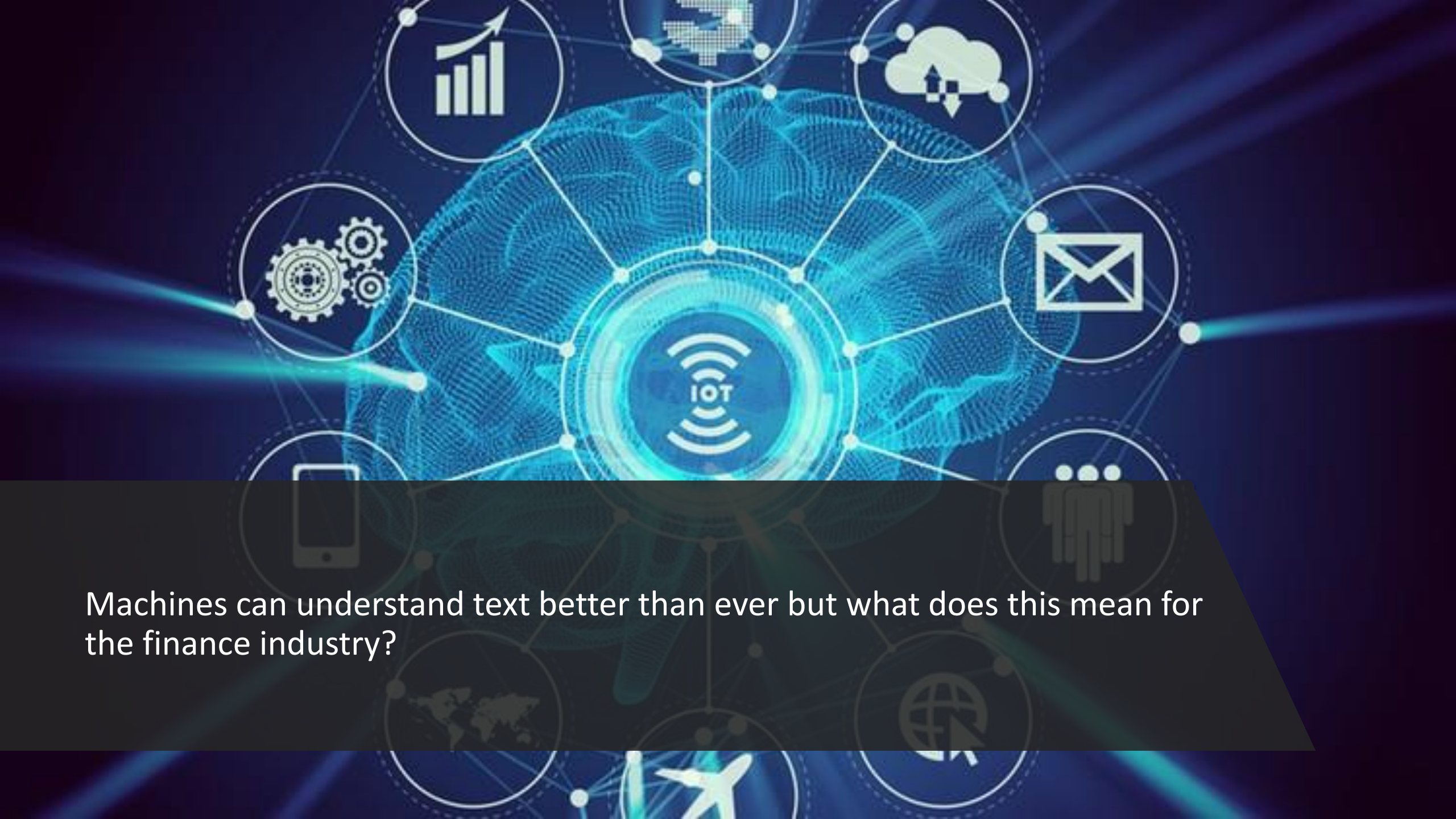
Marcus Reid's Reddit Case-Study



---

Human language is difficult to process, but there is still enough structure to it for machines to find meaning and underlying patterns.





Machines can understand text better than ever but what does this mean for the finance industry?



---

Financial institutions can use text analysis to automate their processes. Using Machine Learning to assist the process can reduce costs and provides a competitive advantage.



---

Why not use Natural Language Processing to help review mortgage and loan contracts? This saves everybody both time and energy.



# Problem Statement

Loan agreements are mostly uncategorized and having people review every contract manually is inefficient. Machine Learning can make life easier and reduce human error.



# Time for a Reddit experiment!

---

I analyzed 754 Reddit posts to see if I can create a machine learning model that could recognize posts from r/anime and ones from r/manga.



# Modeling Process

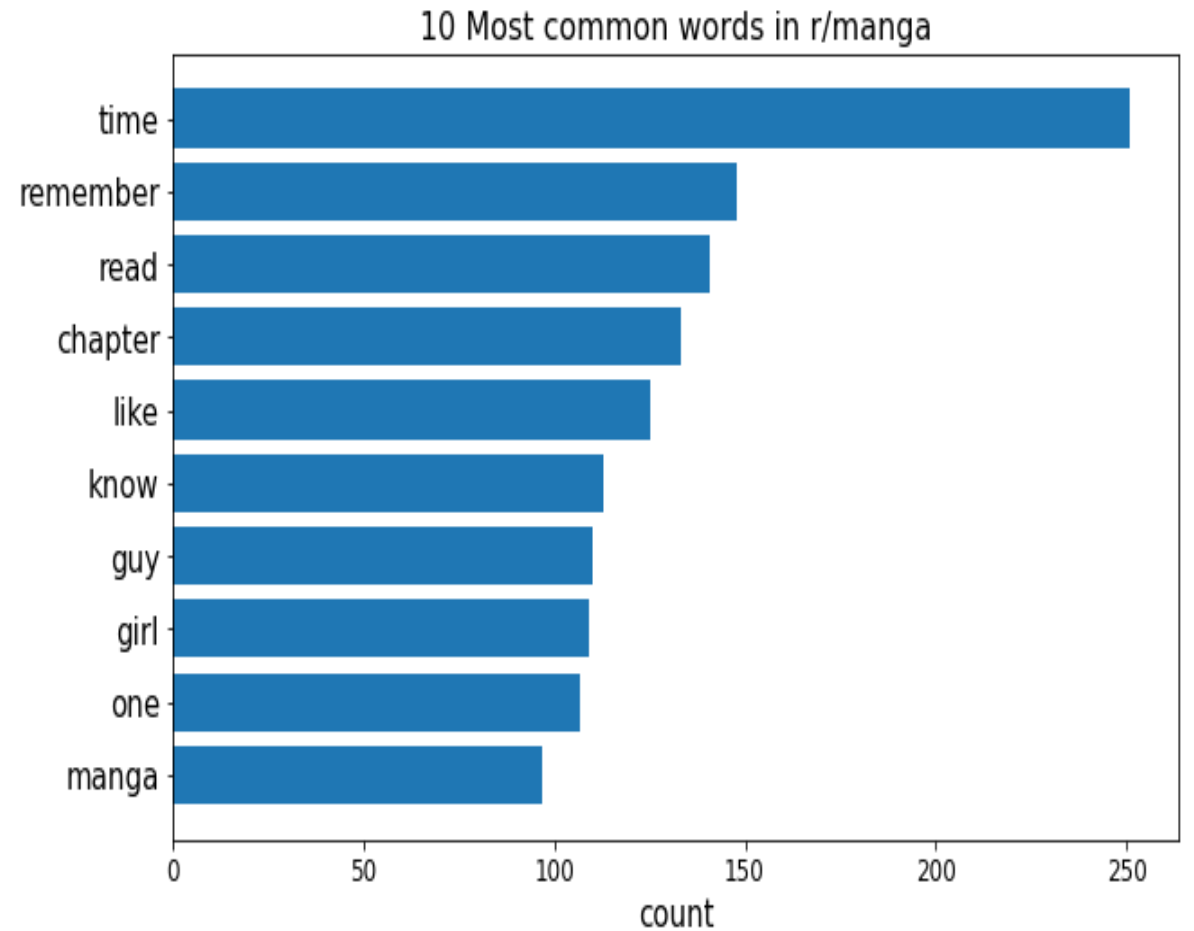
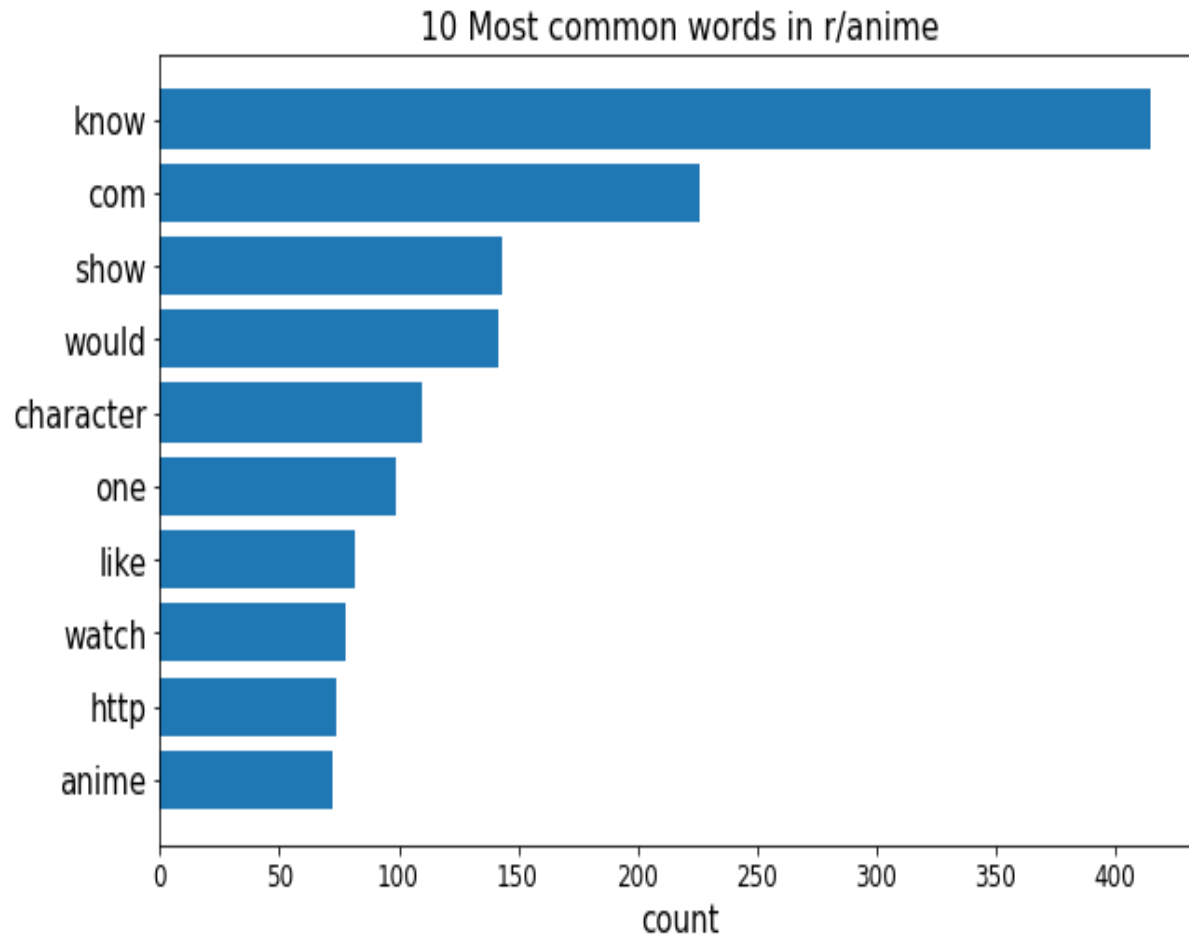
Cleaning and Exploratory Data Analysis

Modeling

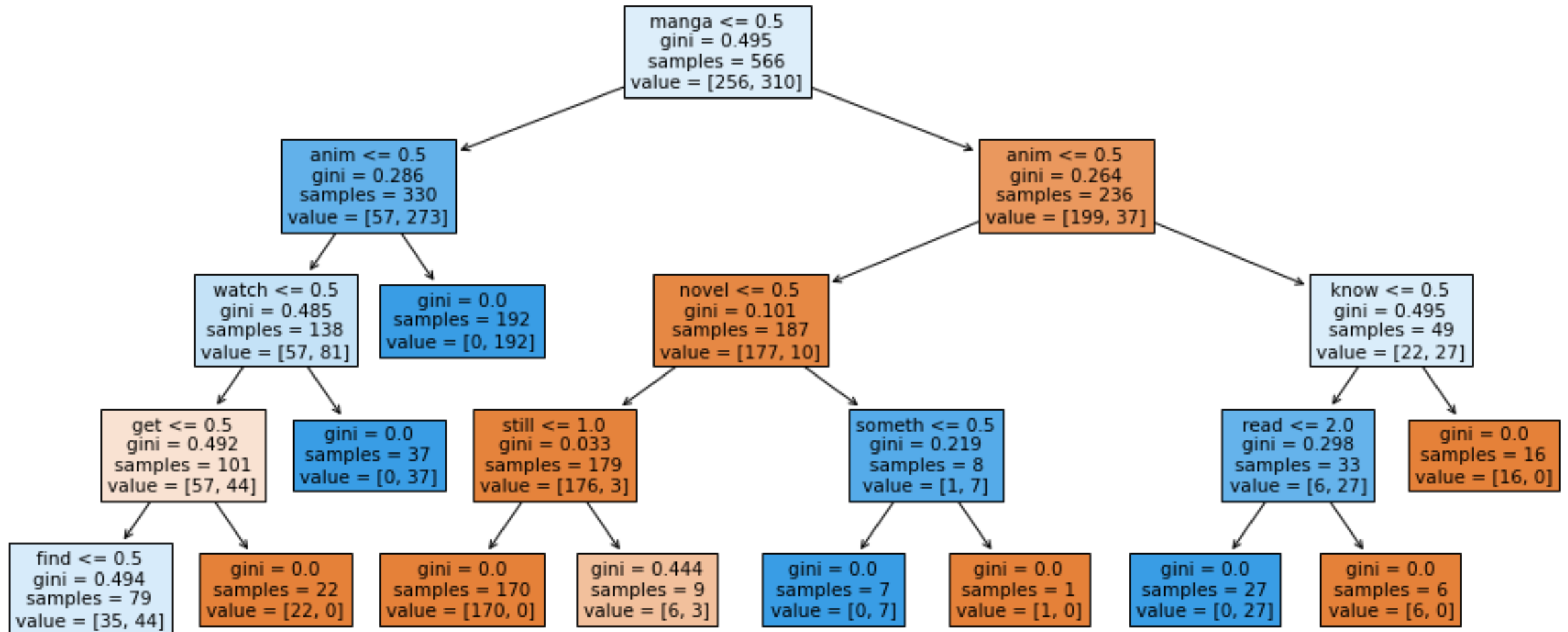
Interpreting Results



# Exploratory Data Analysis

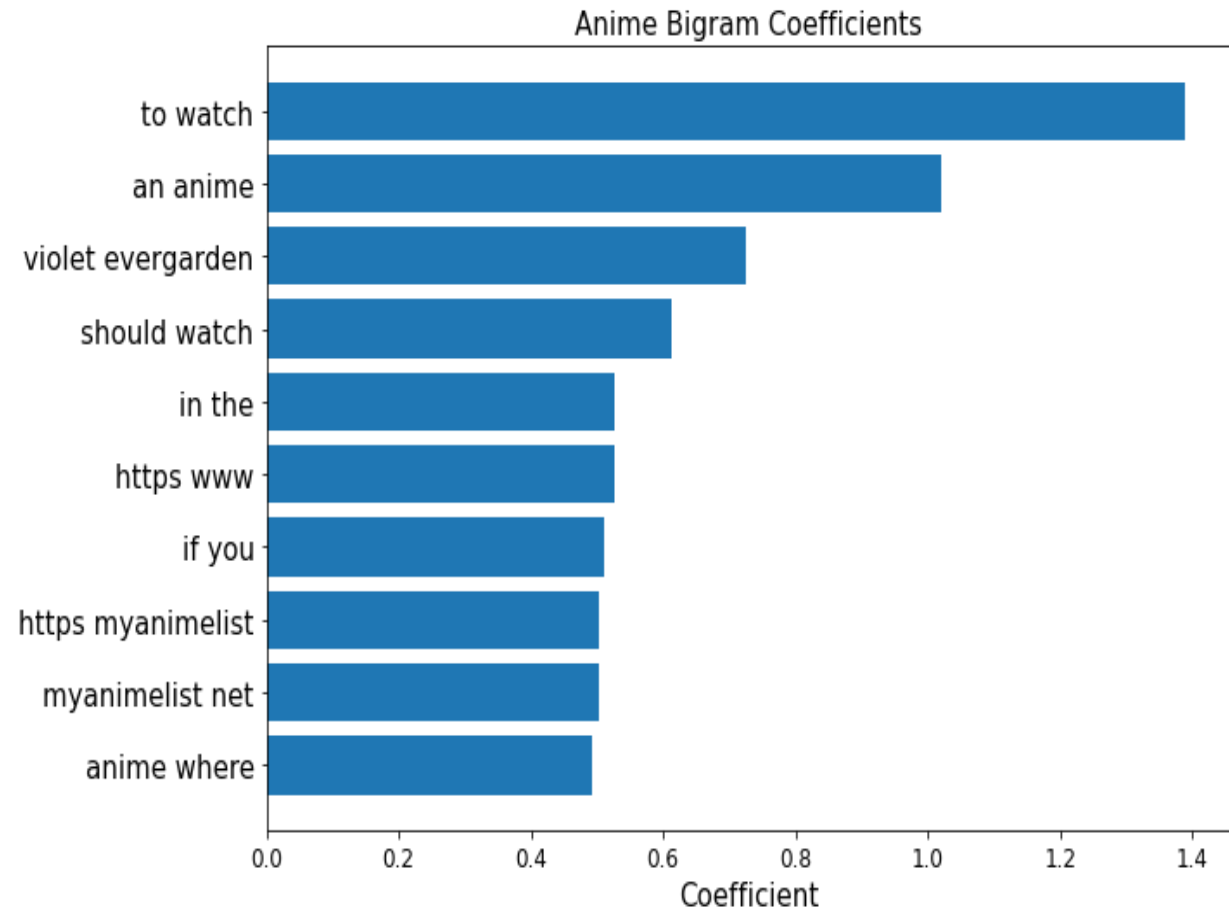
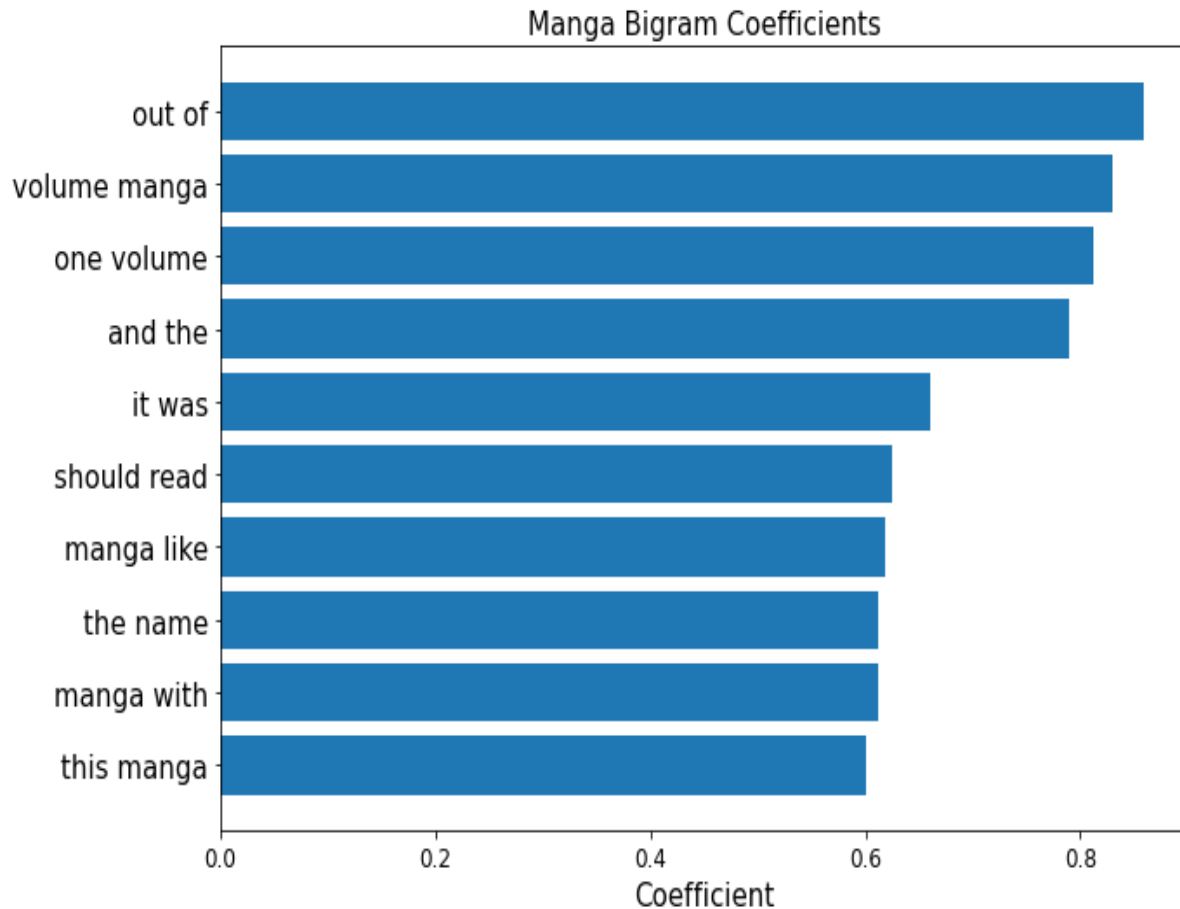


# Modeling



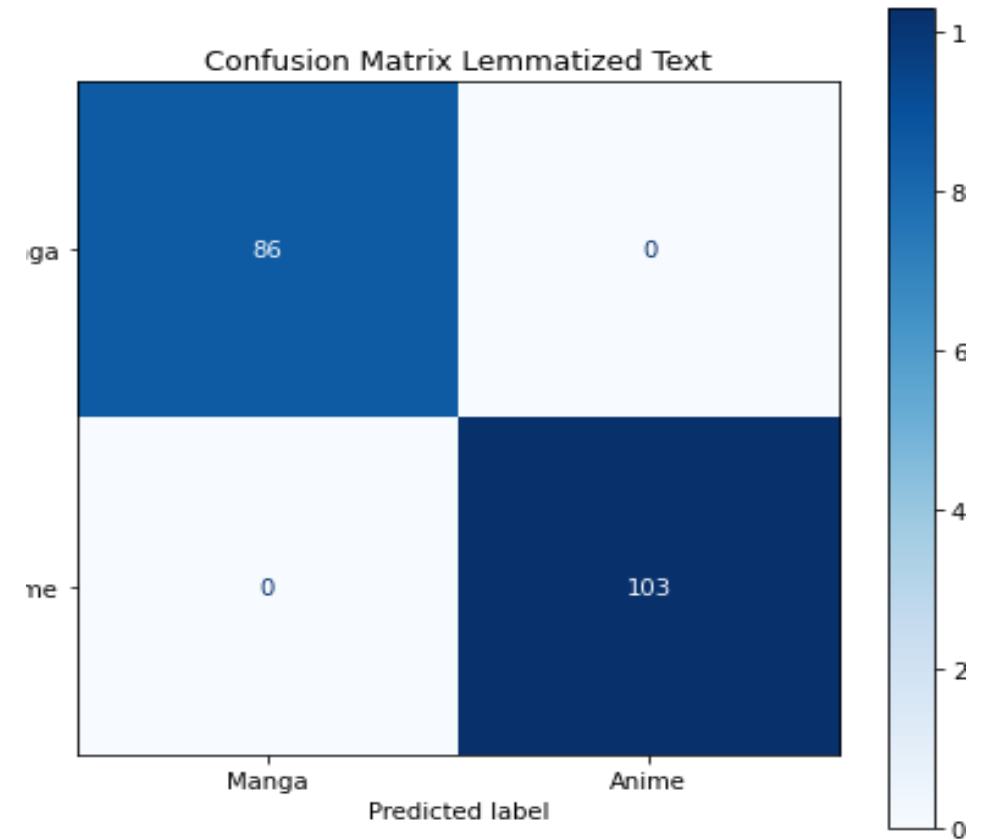


# Modeling



# Results

- {'baggingclassifier\_\_bootstrap\_features': False,  
'baggingclassifier\_\_max\_features': 0.5,  
'baggingclassifier\_\_max\_samples': 0.7,  
'baggingclassifier\_\_n\_estimators': 20,  
'baggingclassifier\_\_warm\_start': True,  
'tfidfvectorizer\_\_max\_df': 1.0,  
'tfidfvectorizer\_\_max\_features': 50000,  
'tfidfvectorizer\_\_ngram\_range': (1, 2)}







## Conclusions and Recommendations

---

A.I. is a powerful tool for classifying text and can save money.

---

Develop custom software to with engineered features for more consistency.

---

An expert needs to work with classification program to ensure accuracy.



# Questions?

---

