An abstract graphic of a tree where the trunk and branches are composed of a network of colored lines (blue, orange, green) and nodes (circles). The tree is positioned on the left side of the slide, with its canopy extending towards the top center.

# U.S. Airline Sentiment Analyses and Prediction Project

**Natural Language  
Processing**

# Business Value



- There are six different airline companies in this dataset and their customers still complaining about some problems with their services/flights. We are going to analyze and making machine learning project for how airline companies could improve themselves with our findings.



**Natural Language  
Processing**

# Business Problem

- In this project, main goal is the analyze and predict airline sentiment(positive,neutral or negative) of flights depend on customer review's **text** with machine learning model. This will help airline companies for future work. Depend on customer's review airline companies could take action about it and improve themselves.

Processing

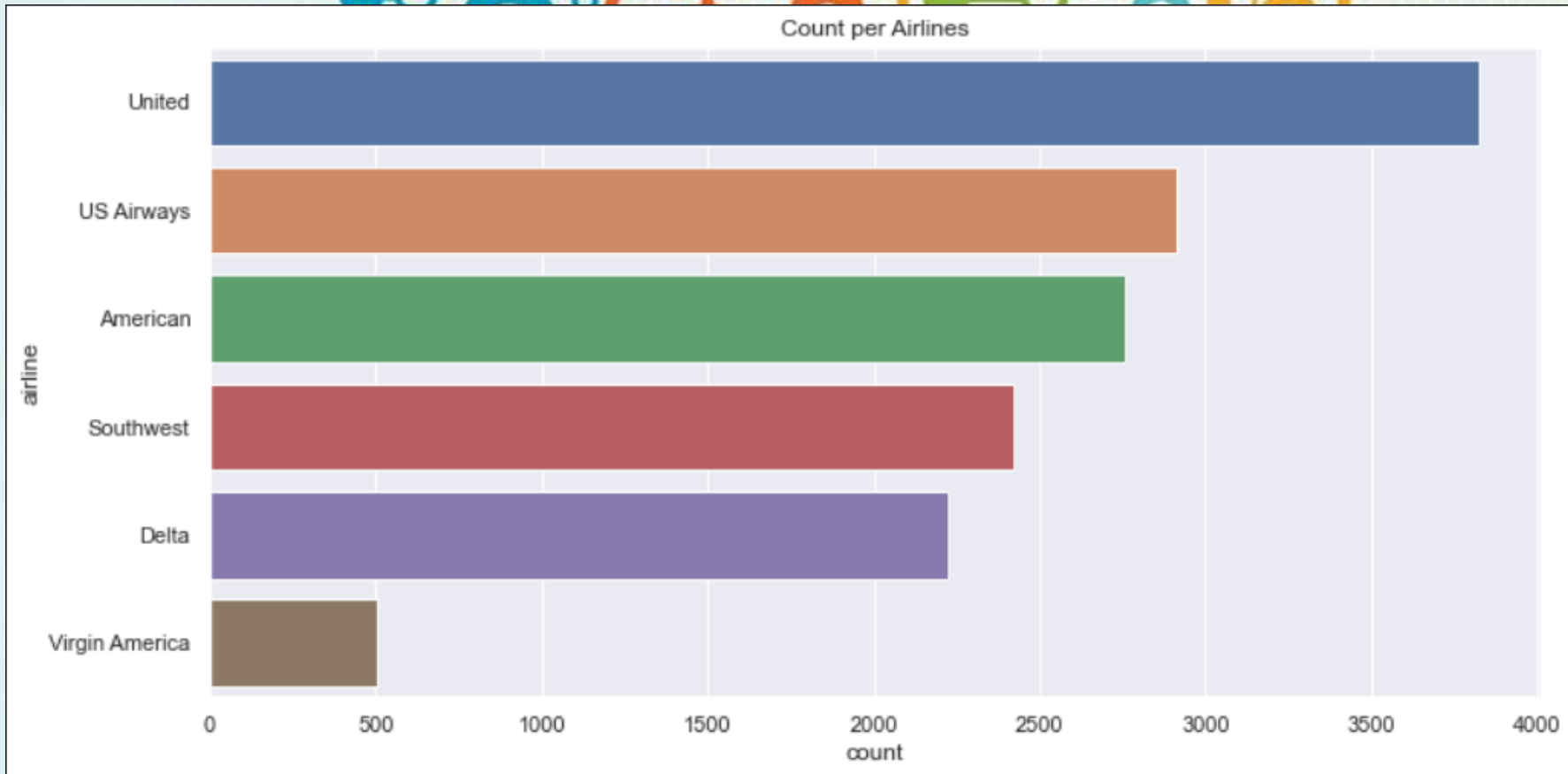


# Data Understanding

1. We have 6 different major U.S. airline companies; United, Us Airways, American , Southwest, Delta and Virgin America.
2. All the tweets scraped at 2015 February.
3. We have three different target category as positive, neutral or negative depend on tweet.

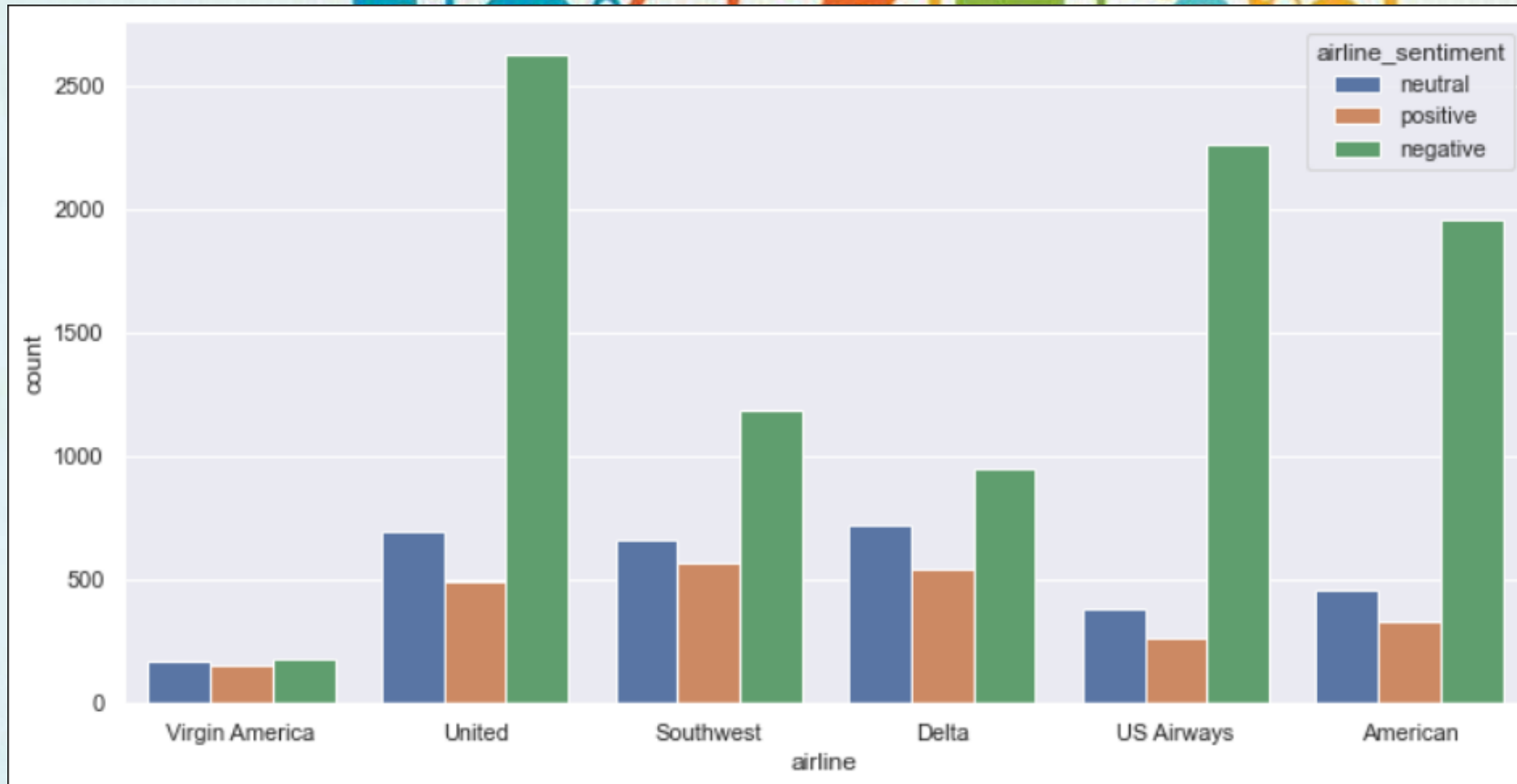
**Natural Language  
Processing**

# Tweet Counts per Airline Companies



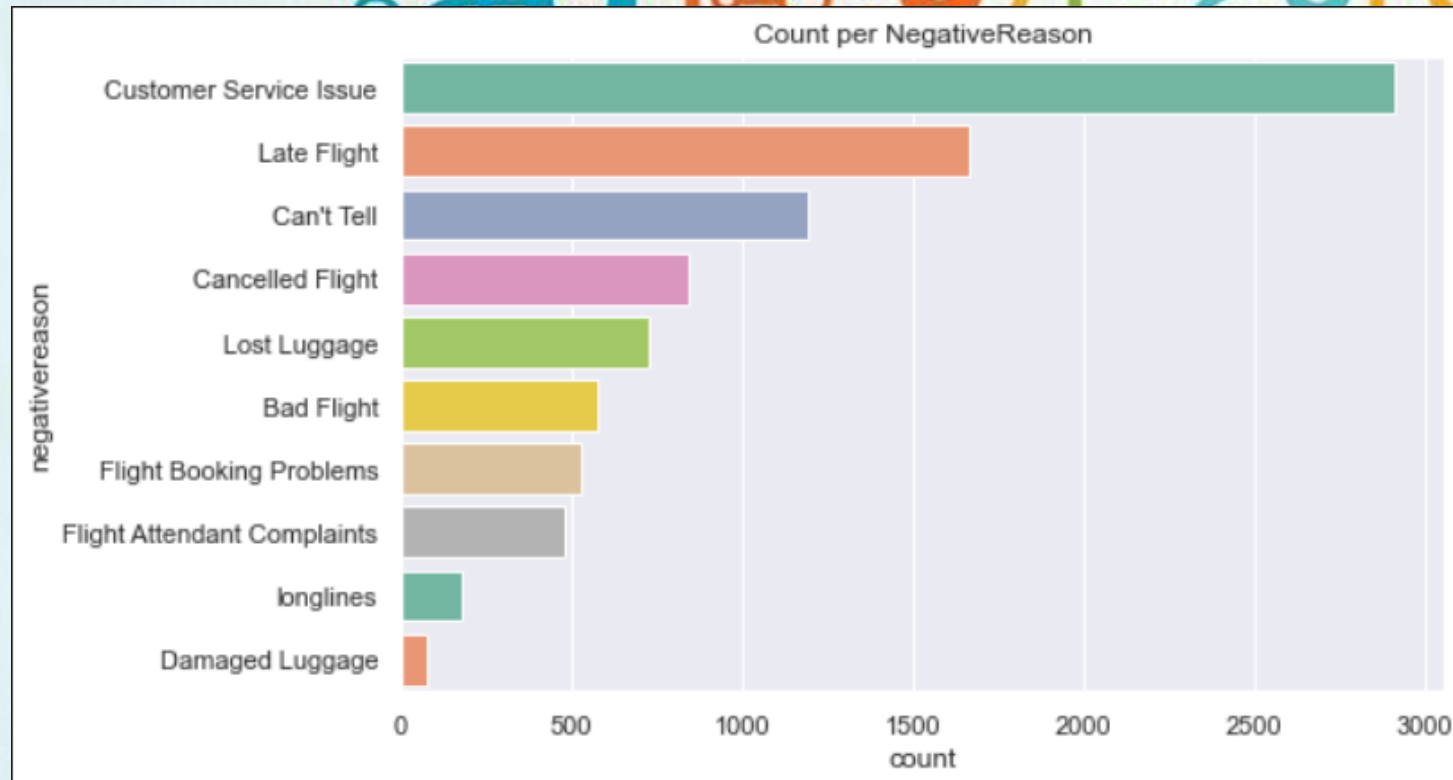
uage

# Sentiment per Airline Companies



In this graph, we can see with greens are negative tweets and most of them are negative. Specially United, Us Airways and American Airlines have the most negative customer reviews.

# Negative Reasons



This visual shows us general negative reasons from customers. Customer service issue and late flight leads for negative reasons.

Language  
ing

**Legend:**

- Customer Service Issue
- Flight Booking Problems
- Can't Tell
- Bad Flight
- Cancelled Flight
- Late Flight
- Flight Attendant Complaints
- Lost Luggage
- Damaged Luggage
- longlines

**Virgin America: 181**

Complaint Category	Frequency
Customer Service Issue	10
Flight Booking Problems	5
Can't Tell	5
Bad Flight	5
Cancelled Flight	5
Late Flight	5
Flight Attendant Complaints	5
Lost Luggage	5
Damaged Luggage	5
longlines	5

**Delta: 955**

Complaint Category	Frequency
Late Flight	250
Customer Service Issue	200
Can't Tell	180
Bad Flight	70
Flight Attendant Complaints	60
Lost Luggage	60
Cancelled Flight	50
Flight Booking Problems	40
longlines	10
Damaged Luggage	10

**United: 2,633**

Complaint Category	Frequency
Customer Service Issue	2500
Late Flight	2200
Can't Tell	1500
Lost Luggage	1000
Bad Flight	800
Cancelled Flight	700
Flight Attendant Complaints	600
Flight Booking Problems	500
longlines	200
Damaged Luggage	100

**US Airways: 2,263**

Complaint Category	Frequency
Customer Service Issue	2100
Late Flight	1800
Can't Tell	1200
Cancelled Flight	1000
Lost Luggage	800
Flight Attendant Complaints	700
Flight Booking Problems	600
Bad Flight	500
longlines	300
Damaged Luggage	100

**Southwest: 1,186**

Complaint Category	Frequency
Customer Service Issue	1000
Cancelled Flight	500
Can't Tell	400
Late Flight	300
Lost Luggage	200
Bad Flight	150
Flight Booking Problems	100
Flight Attendant Complaints	50
longlines	50
Damaged Luggage	20

**American: 1,960**

Complaint Category	Frequency
Customer Service Issue	1700
Late Flight	600
Cancelled Flight	500
Can't Tell	400
Lost Luggage	300
Flight Booking Problems	200
Bad Flight	150
Flight Attendant Complaints	100
longlines	50
Damaged Luggage	20



# Conclusion

In this project tried different machine learning models to find best result. Some of basic models name are ; Logistic Regression, Random Forest, Gradient Boosting. But improved model called neural network model gives us best result as **92%** accuracy on dataset.

What this number mean is model will predict 92% true (positive,neutral or negative) on given tweet. For example;

We will give our model as tweet "didn't enjoy flight". Model result will be 'Negative'.

## Processing

# Future Step



- We can work on detail text and improve our model.
- For example we still get little bit of false positive, neutral or negatives. We would work why is that and what we could do more on our model.
- Also gather more tweets about these airline companies.



**Natural Language  
Processing**



Thank  
You

By Huseyin Caglar

[huseyincaglar@outlook.com](mailto:huseyincaglar@outlook.com)