## PERFORMANCE COMPARISON BETWEEN SUPPORT VECTOR MACHINE, NAÏVE BAYES CLASSIFIER, SEQUENTIAL MODEL, AND CONVOLUTIONAL NEURAL NETWORK FOR TEXT DATASET SENTIMENT ANALYSIS

## Aamod Khatiwada, Nitisha Patange, Humberto Pinon

### **Abstract**

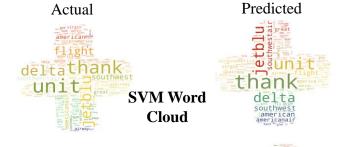
project compares the classification performances among the Naïve Bayes Classifier, Support Vector Machine (SVM), Sequential Model, and Convolutional Neural Network (CNN). The airline review twitter dataset from Kaggle was used, which consists of Positive, Negative, and Neutral as the classification labels. The models were trained using the same dataset and the four different algorithms. The best classification accuracy obtained from all four models were used for the final comparison. The visualization tools like graphs and word clouds are used to support the results. The project concludes that the deep learning models are more accurate on the classification task for the selected dataset.

## **Primary Objective**

the compare text sentiment classification accuracy of SVM, Naïve Bayes, sequential, and CNN for an Airline review twitter dataset.

#### **Dataset Introduction**

- Dataset: Twitter US Airline Sentiment
- Dataset was obtained from Kaggle.
- The tweets were collected according to 3 classes: positive, negative and neutral.
- Total number of Tweets: 3000
- 2400 for training (80%).
- 600 for testing (20%).
- Positive: 1000 • Negative: 1000
- Neutral: 1000





Naïve Bayes **Word Cloud**  thank delta

## **Convolutional Neural** Network

• 1D convolutional layer and 1 pooling layer

**Sequential Model** 

• Number of epochs (number

cycle through the data): 6

of times the model will

• 1 embedding layer and 1

• Activation Function:

dense layer

Sigmoid

- Activation Function: Sigmoid
- Number of epochs (number of times the model will cycle through the data): 6
- 128 kernels (kernel size = 5)

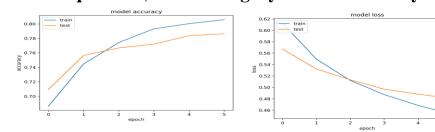
## SVM classifier for Gamma=0.7, C=1.5

Accuracy: 69.16666666666667%										
	precision	recall	f1-score	support						
negative	0.72	0.70	0.71	213						
neutral	0.57	0.66	0.62	185						
positive	0.80	0.70	0.75	202						
accuracy			0.69	600						
macro avg	0.70	0.69	0.69	600						
weighted avg	0.70	0.69	0.69	600						

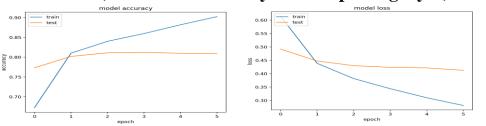
# Naïve Bayes Classifier

Accuracy: 67.333333333333333333333333333333333333								
	precision	recall	f1-score	support				
negati	ve 0.66	0.78	0.72	213				
neutr	al 0.62	0.48	0.54	185				
positi	ve 0.73	0.74	0.74	202				
accura	су		0.67	600				
macro a	vg 0.67	0.67	0.66	600				
weighted a	vg 0.67	0.67	0.67	600				

## Sequential (1 embedding layer and 1 dense layer)



## CNN (1D convolutional layer and 1 pooling layer)



#### Conclusion

- Naïve Bayes and SVM: Almost equal accuracy.
- Sequential model and CNN: Almost equal accuracy.
- Higher classification Accuracy seen on deep learning models with CNN leading among the two.
- Most inaccurate classification in general: **Neutral Sentiment Texts**