



NLP Emotion Presentation

DHANI POZNIAK



Overview

- Documentation is a form of communication
- Instantaneous
- Messenger
- Texting
- RealTime chat
- Chat bots





Business Understanding

- NLP used to improve products by reviews
- Improve well-being
- Self-evaluation
- Guide to assist in therapy assessment
- Promote positive behaviour patterns

Data Understanding

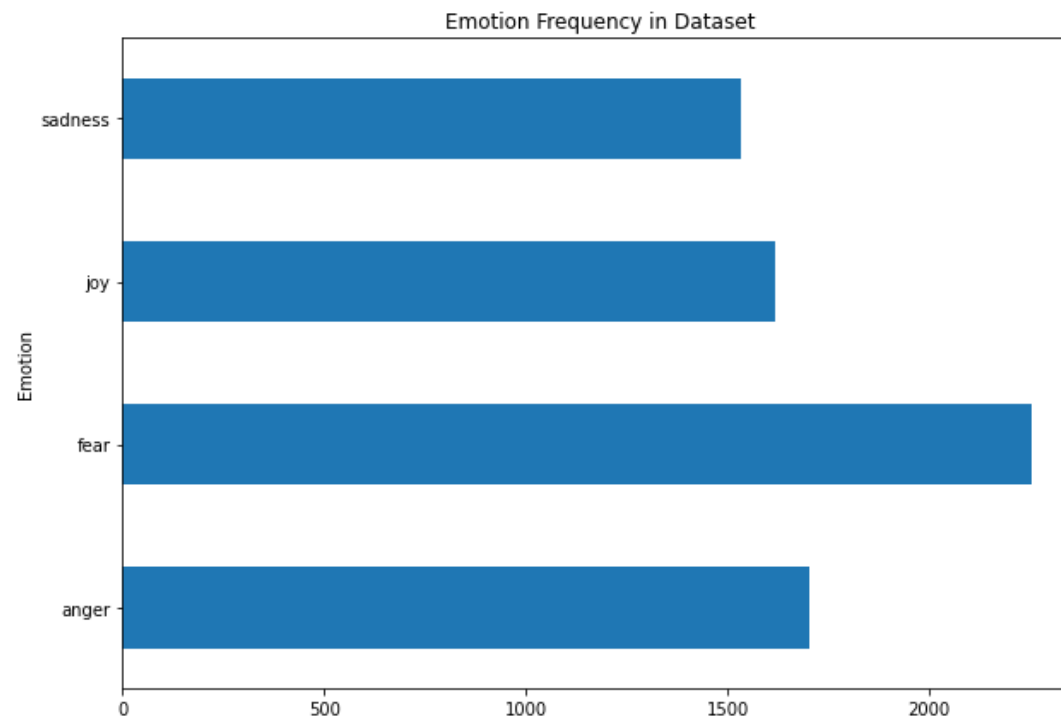
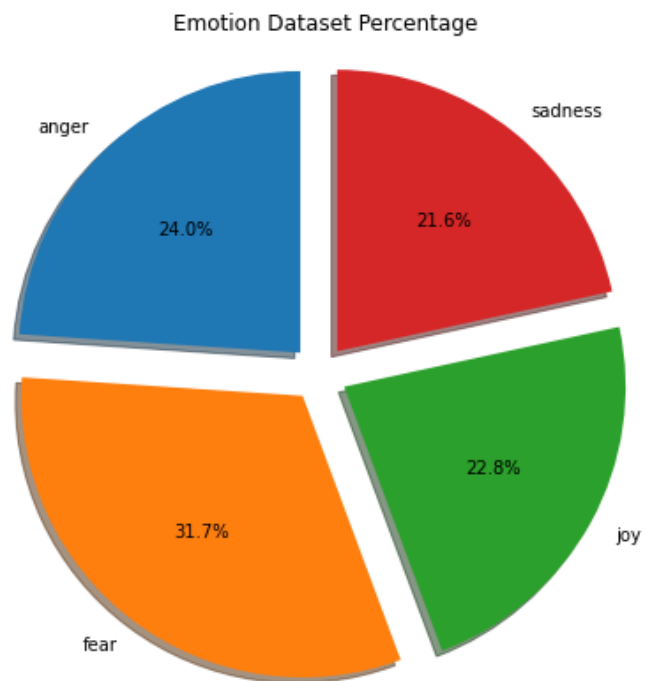
- Dataset:


A. Tripathi, "Emotion Classification NLP",
Kaggle.com, 2021. [Online].
Available: <https://www.kaggle.com/datasets/anjaneyatripathi/emotion-classification-nlp>.
[Accessed: 16- Jul- 2022].

- 7102 Rows, 2 Columns
- Joy, Sadness, Anger, Fear

Data Preparation

- Initial analysis shows a balanced dataset



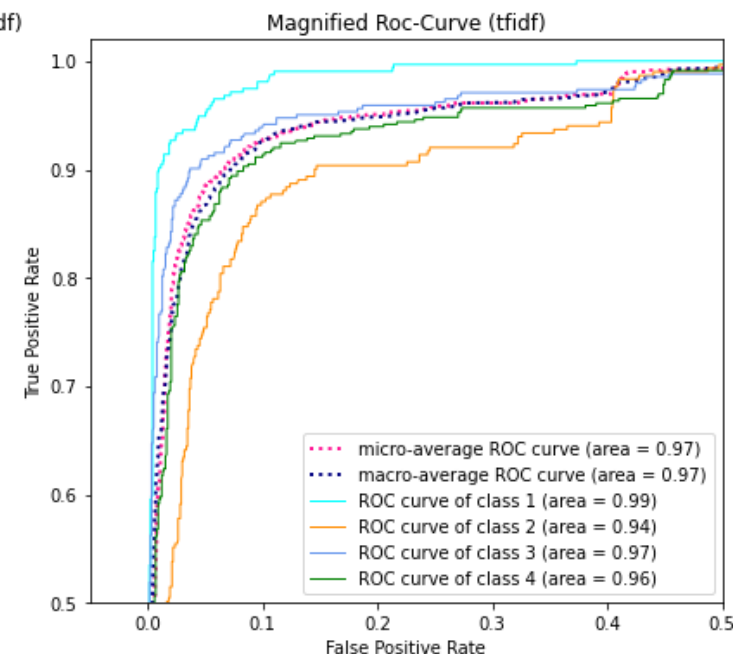
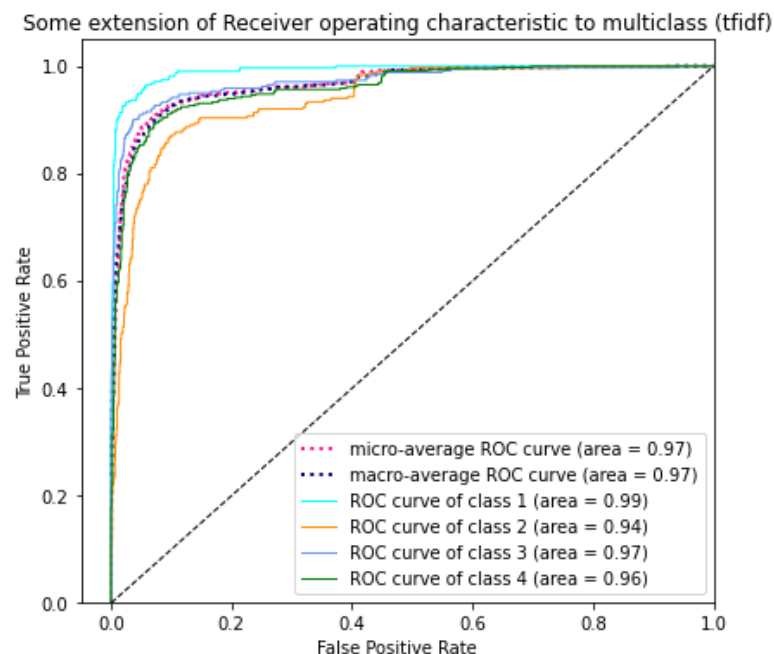
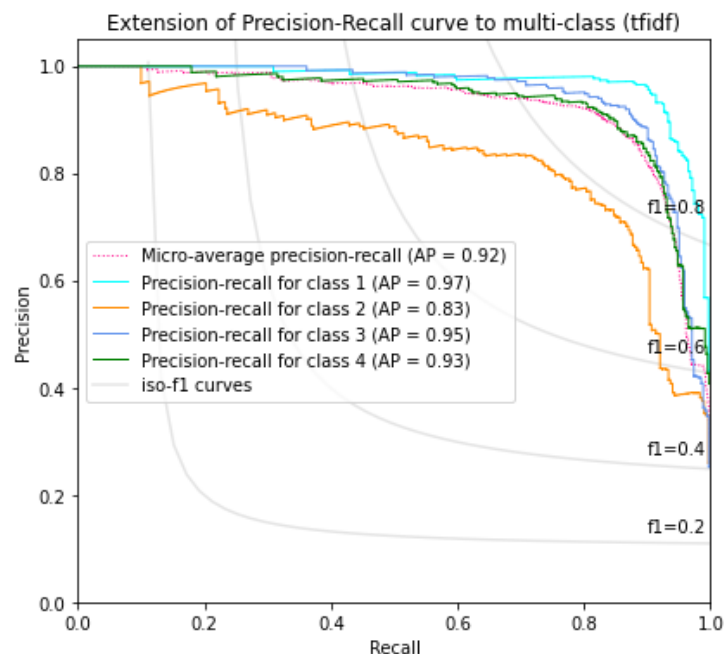


Text Preparation

- handle encoding
- handle extraneous and international characters
- handle symbols/ emojis
- handle metadata and embedded information
- handle repetitions (such multiple spaces or newlines)
- Stopwords
- Stemming and lemmatization
- Important phase to be reproducible

Baseline Model

- Used count and tf-idf vectorizer with LinearSVC
- Accuracy: 84.34% with tf-idf/ LinearSVC



Feature Engineering

- Text Features
- PCA
- Topic Words
- Decrease in accuracy score



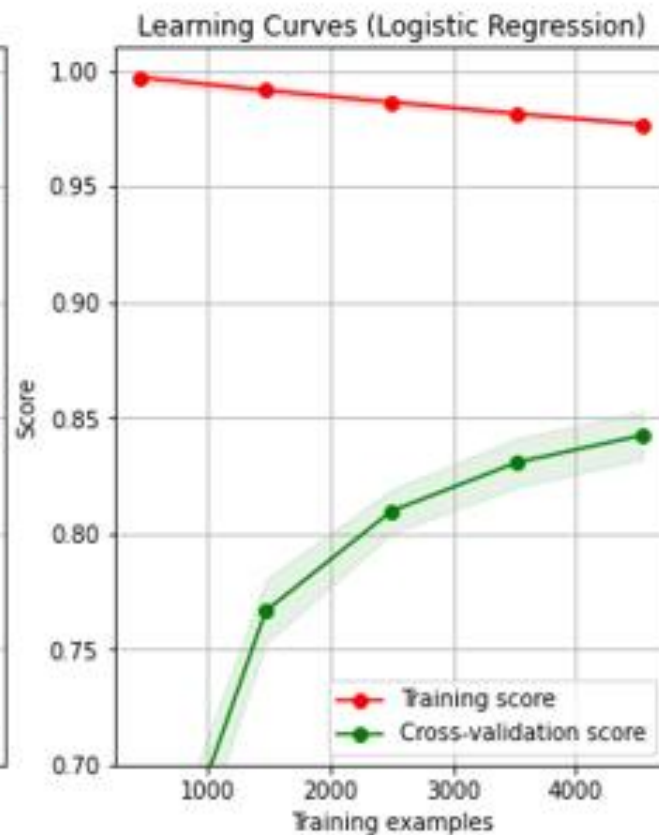
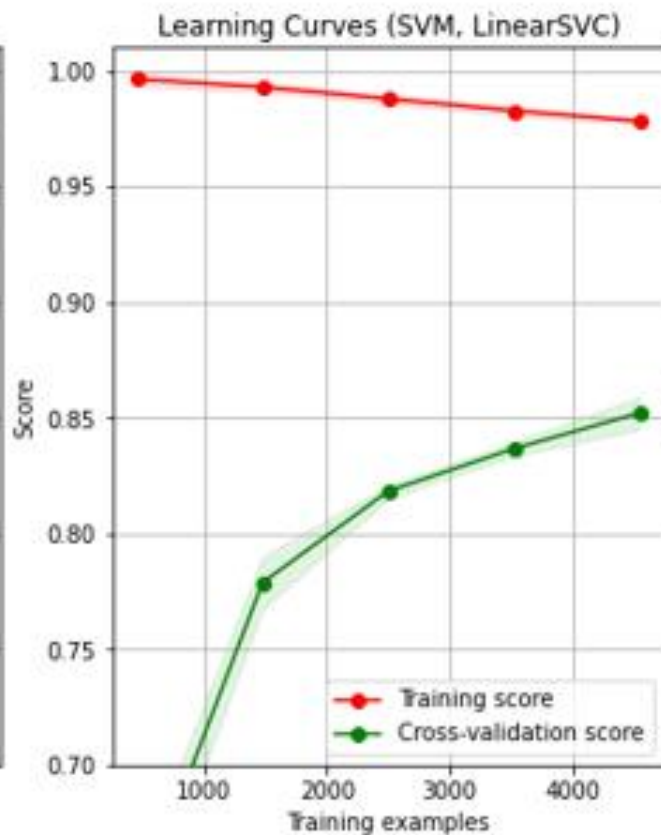
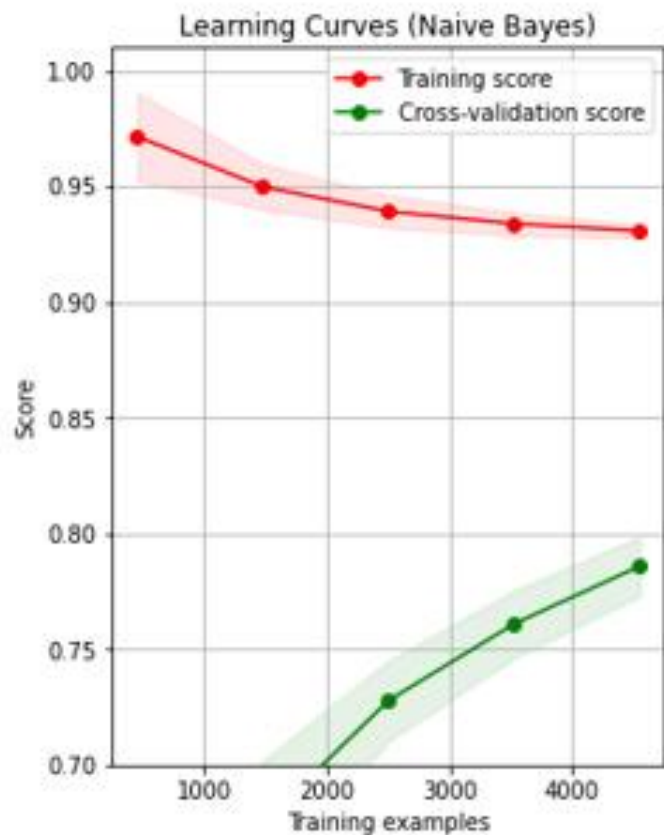
Model Comparison- Phase 01

	Count Vectors	WordLevel TF-IDF	N-Gram Vectors	CharLevel Vectors
Naive Bayes	0.830	0.809	0.395	0.695
Logistic Regression	0.843	0.848	0.395	0.780
Support Vector Machine	0.836	0.855	0.388	0.803
Random Forest	0.835	0.828	0.391	0.764
Gradient Boosting	0.814	0.811	0.384	0.788

- Chose one of each machine algorithm with highest accuracy score to move onto phase 02

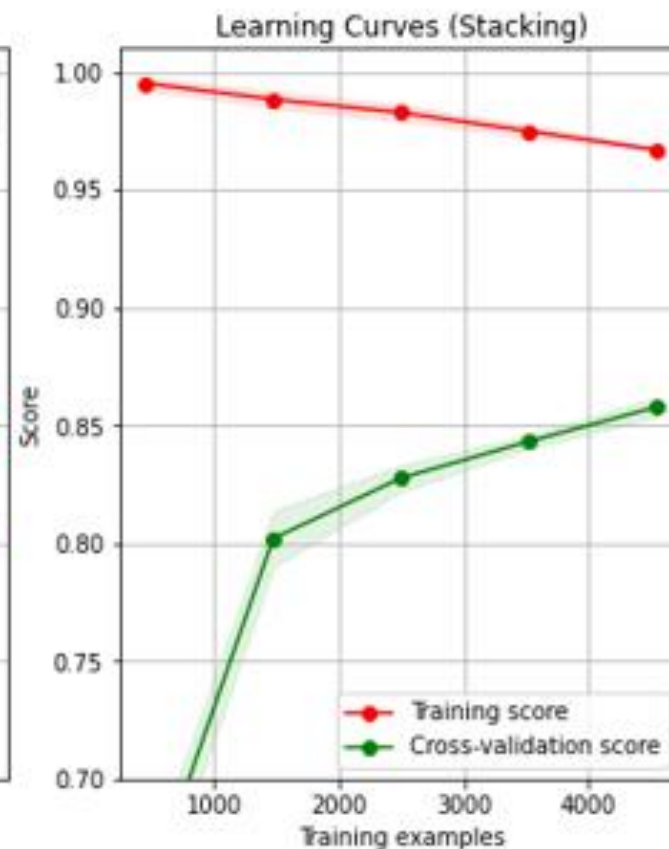
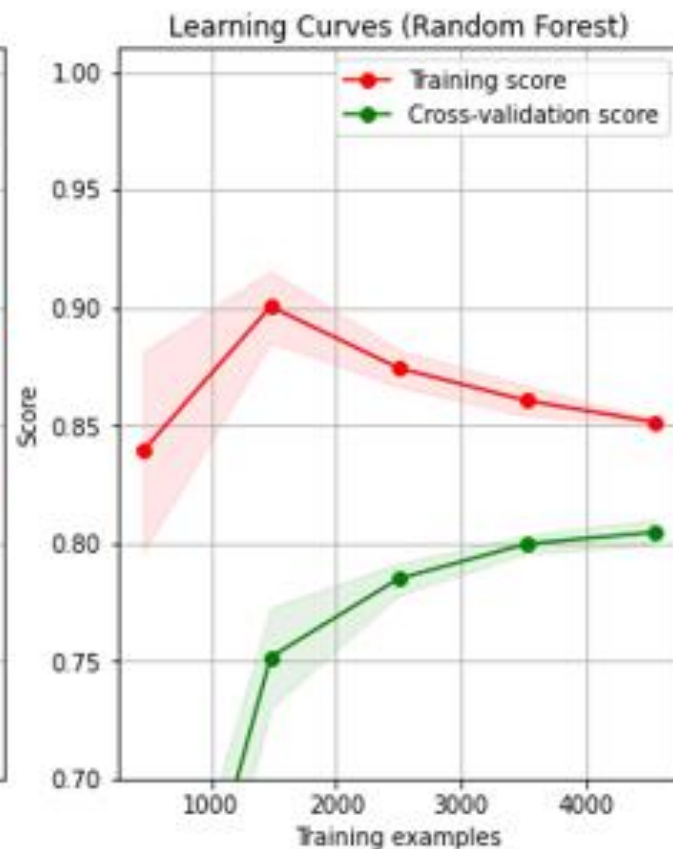
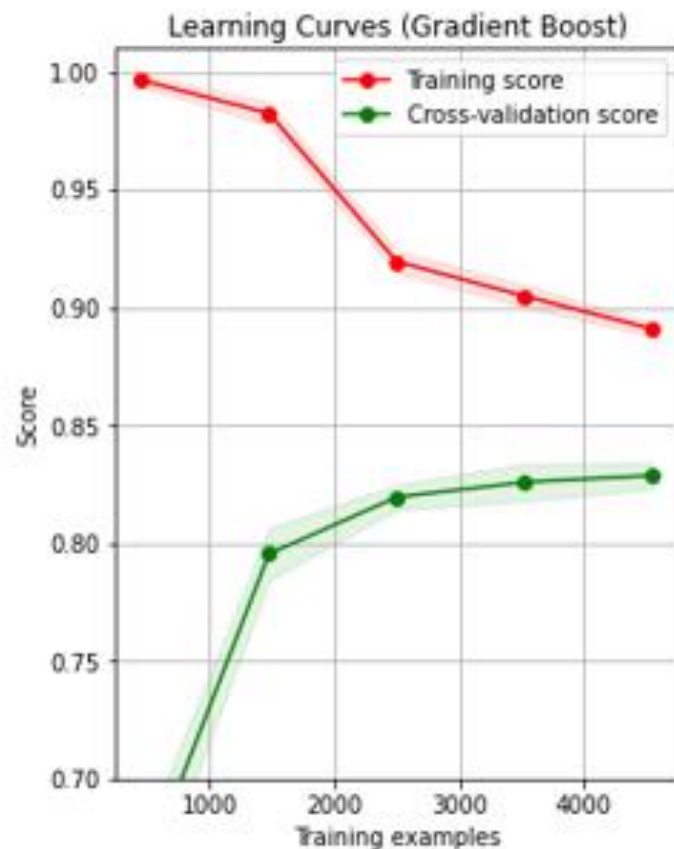
Learning Curves (GNB, SVC, LogReg)

- Baseline Accuracy: 0.843

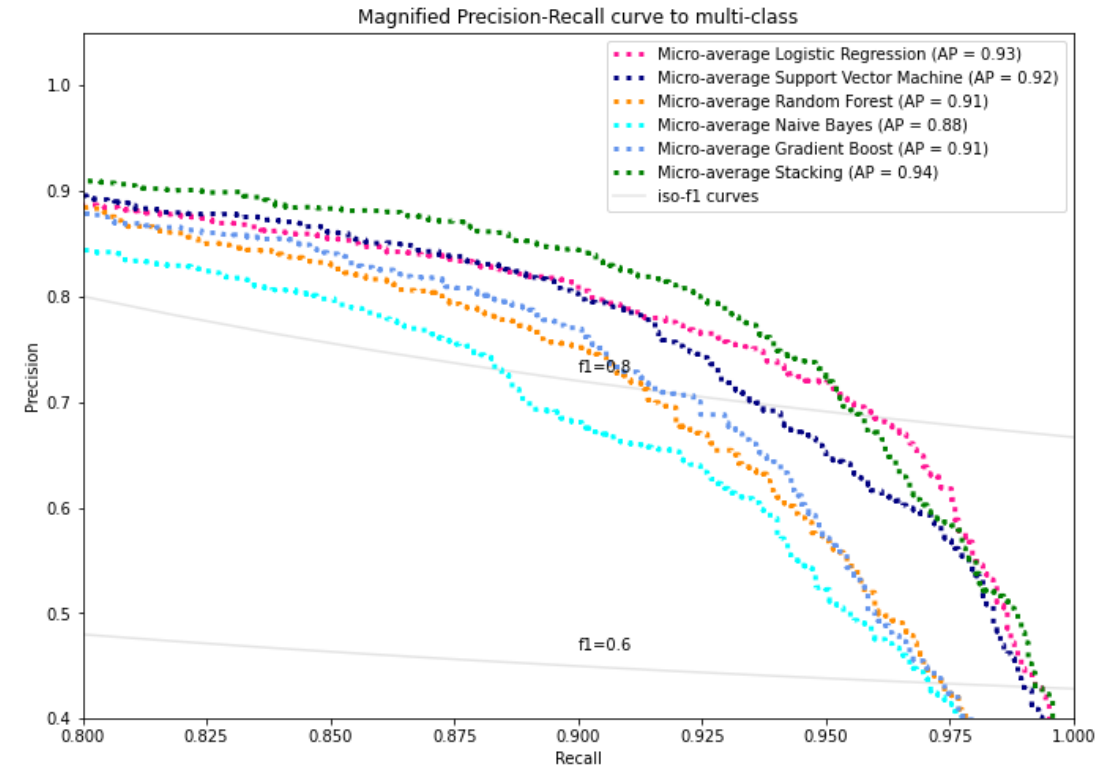
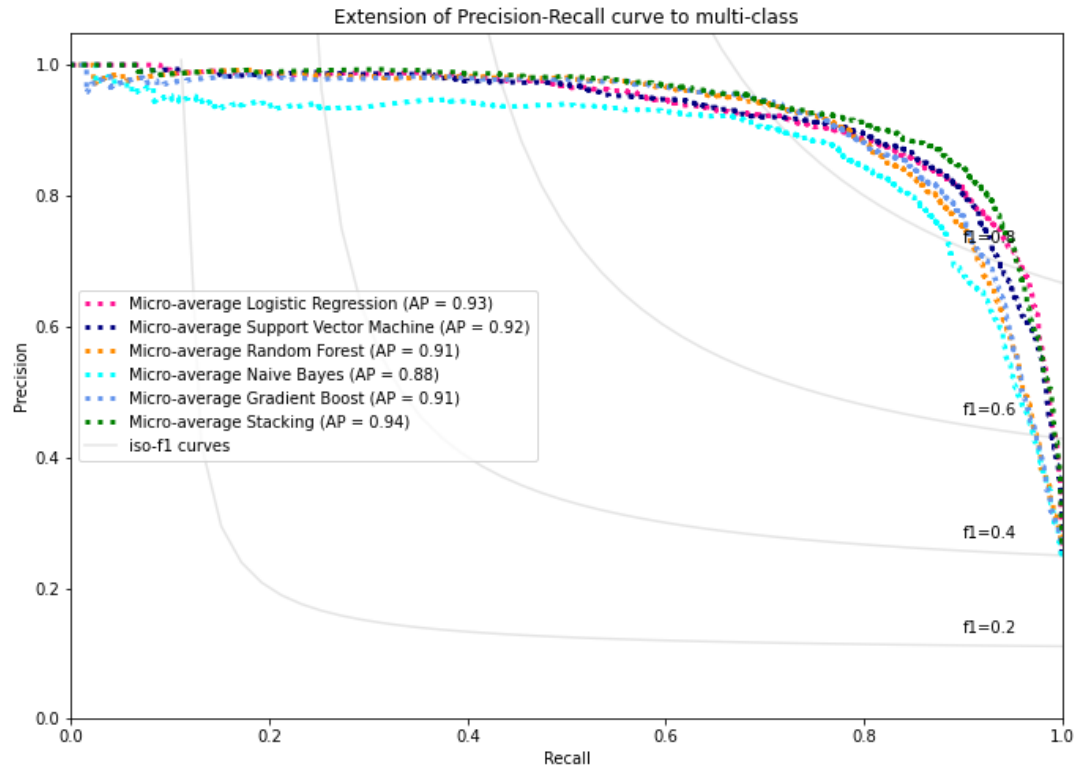


Learning Curves (GradBst, RFC, Stacking)

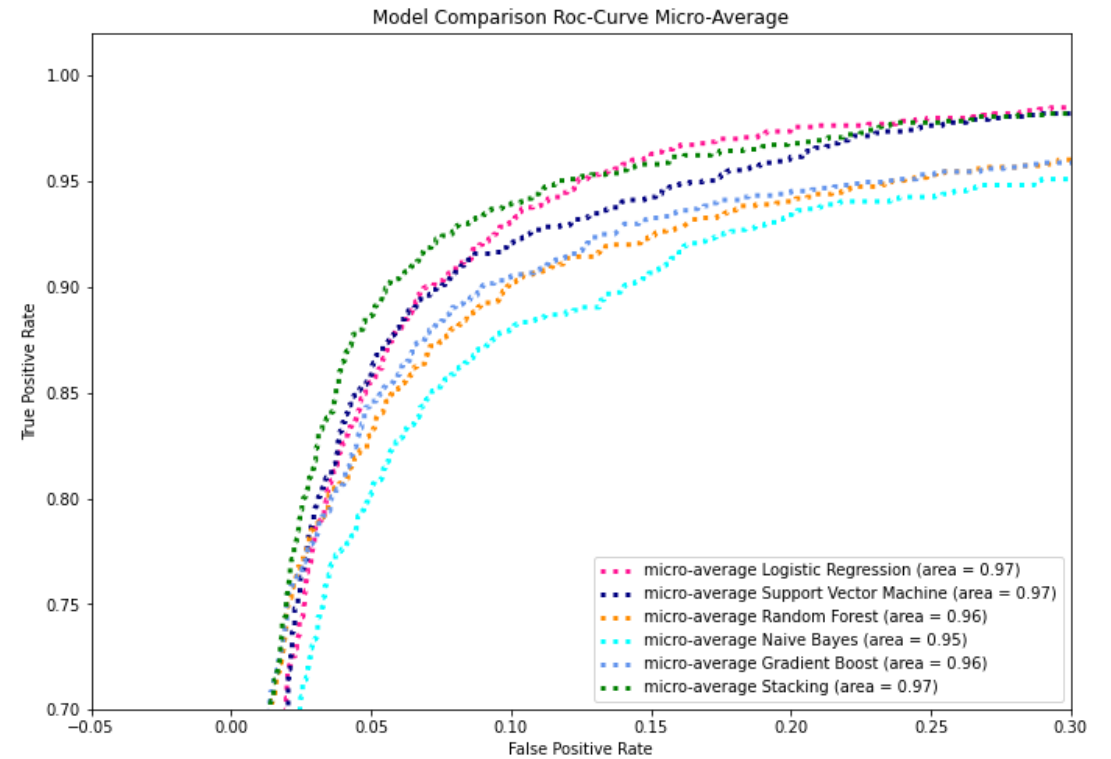
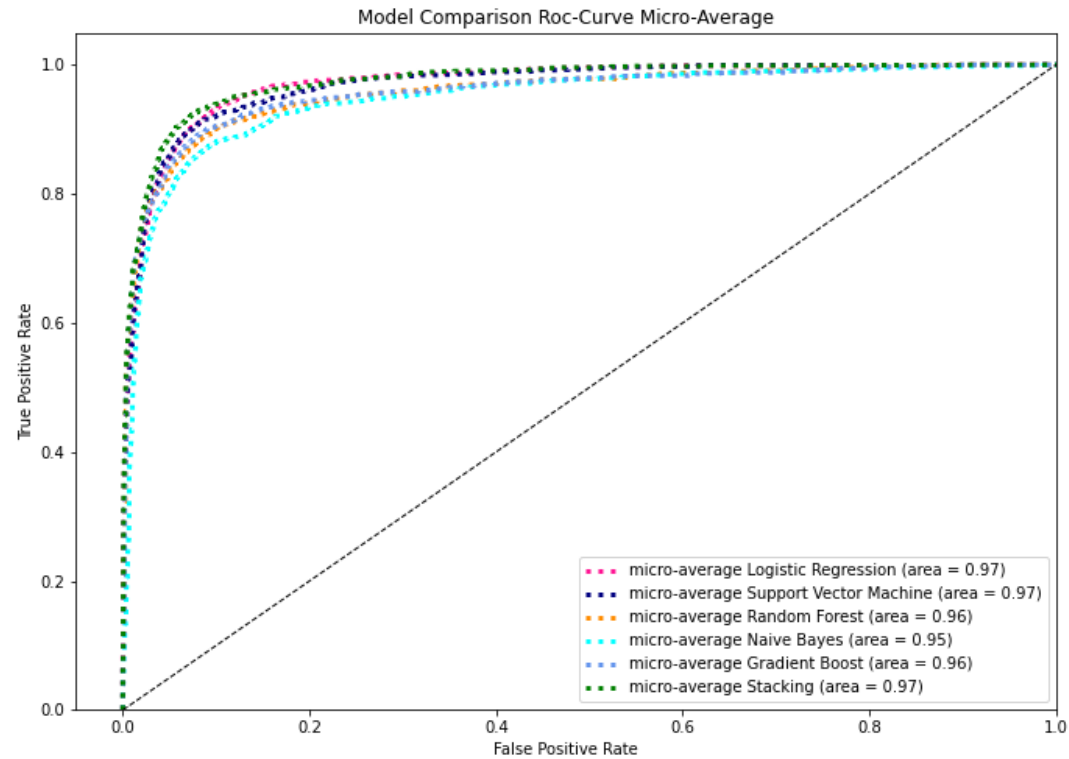
- Baseline Accuracy: 0.843



Precision/ Recall Curve



Roc-Curve

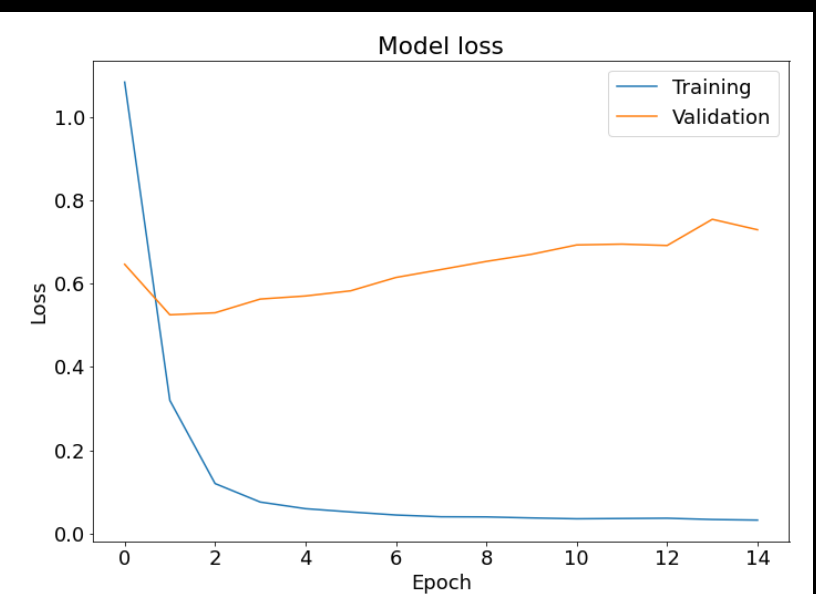
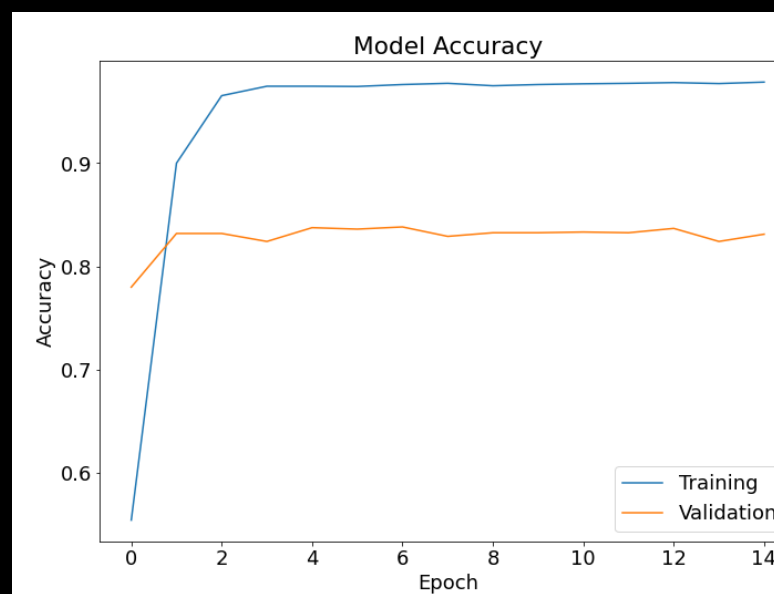


Bonus: Neural Network

- 3 Hidden Layers
- [100, 25, 15] neurons
- Output Layer: 4 neurons
- 15 epochs

Confusion Matrix (Neural Network)

	joy	sadness	anger	fear
joy	293	24	16	22
sadness	6	216	28	33
anger	6	29	271	12
fear	18	47	12	388
Predicted label	True label			



Model Comparison- Phase 02 Hyper-Parameters

- Baseline Accuracy: 0.843

	accuracy	precision	recall	f1-score	roc_auc	runtime (ms)	meanAcc
logisticRegression tf_idfVect	0.849	0.853	0.845	0.849	0.966	1050	0.843
supportVectorMachine tf_idfVect	0.860	0.860	0.857	0.859	0.970	243	0.856
randomForestClassifier countVect	0.839	0.873	0.826	0.840	0.971	567	0.821
naiveBayes countVect	0.830	0.833	0.824	0.827	0.950	4	0.798
gradientBoosting countVect	0.854	0.878	0.845	0.857	0.974	4394	0.842
stacking countVect	0.874	0.878	0.870	0.873	0.974	19341	0.854
stacking tfidfVect	0.872	0.873	0.869	0.871	0.974	12787	0.863
neuralNetwork countVect	0.820	0.822	0.816	0.819	-	308	0.622



Superior Model



Model Accuracy > Baseline Model

Superior Model

- Support Vector Machine (LinearSVC)
- Evaluation speed: 243 ms
- Implementing the chosen legitimate model, it is able to predict with 86% confidence matching emotion to the related text as either:
 - Joy
 - Sadness
 - Anger
 - Fear

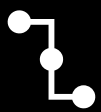
Confusion Matrix (SVC)

Predicted label \ True label	joy	sadness	anger	fear
joy	296	9	6	7
sadness	4	245	29	31
anger	6	23	278	14
fear	17	39	14	403

Deployment

- Integrate with existing technologies
- Messenger
- Texting
- RealTime chat
- Chat bots
- Improve person's well-being

Contact



LinkedIn

<https://www.linkedin.com/in/dhani-pozniak-34801a196/>



Github

<https://github.com/pozengineer>



Email

pozengineer@hotmail.com



Phone

0412 966 933



Thank You

Any Questions ?