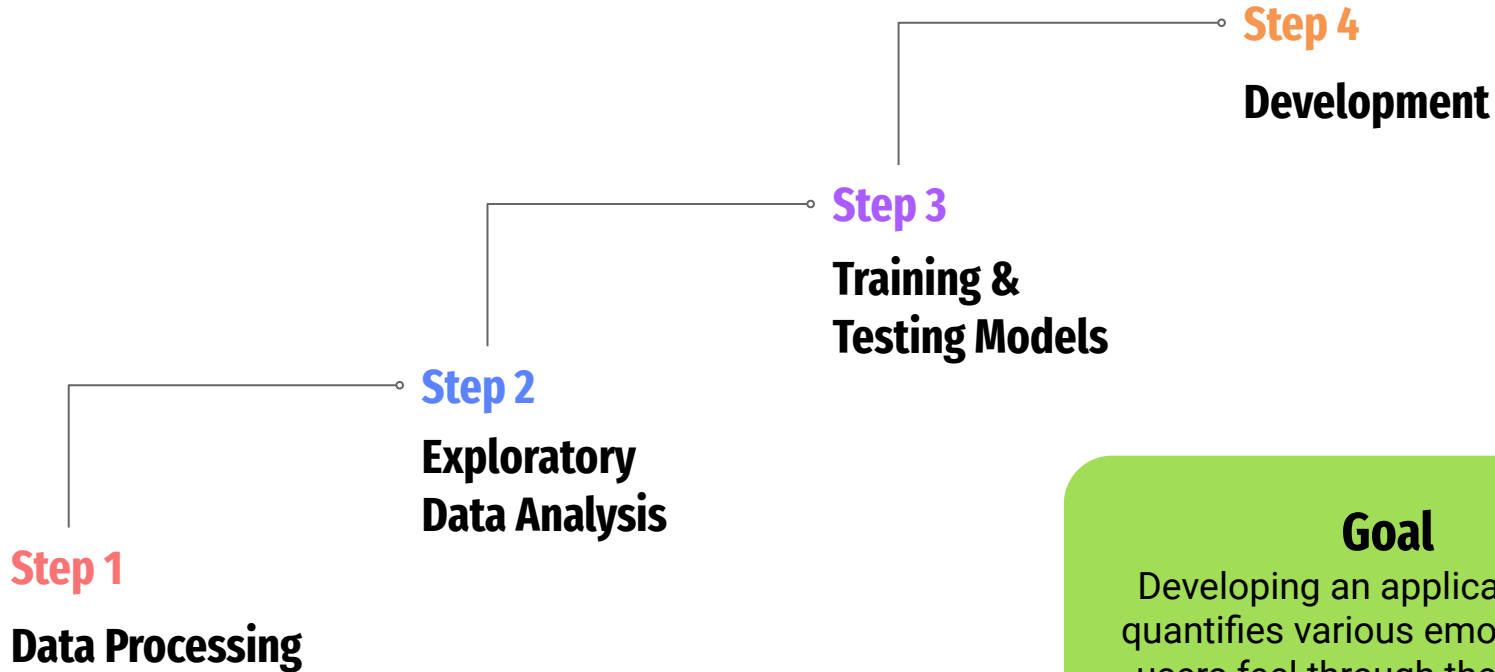


Emotion Recognition Tool for Tweets

Capstone Project Presentation

Harshit Bhavnani

Methodology

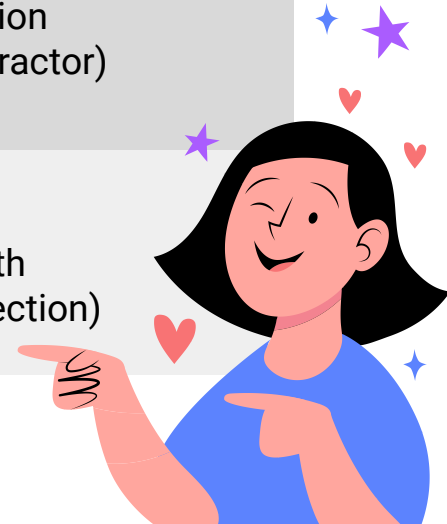


Data

ID	Tweet	Affect Dimension	Intensity Score
2017-En-30692	Positive research show salesperson score top g...	joy	0.274
2017-En-11102	passed away early morning fast furious styled ...	anger	0.354
2017-En-41401	If Troyler die Im gonna die	sadness	0.798
2017-En-21664	terrorism booming industry Pak govt oblivious ...	fear	0.625

Methodology

ETL Process	
Cleaning	Preparation
Tagged Accounts & Links	Feature Extraction (Textblob Conll Extractor)
Empty/Duplicate Tweets and NA Values	
Punctuation Marks & Special Characters	Embedding (Word2vec with Google News Collection)
Stopwords (Standard English from NLTK)	



Experimental Study

Linear Regression

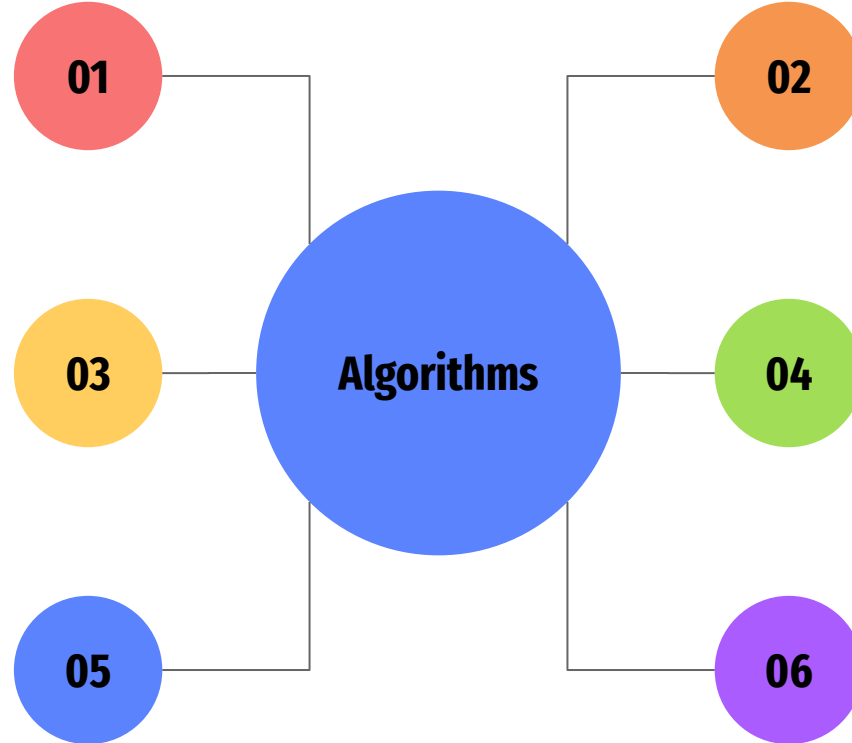
A linear approach for modelling the relationship between a scalar response and one or more explanatory variables

Random Forest

An ensemble learning method that operates by constructing a multitude of decision trees at training time.

Decision Tree

Breaks down a dataset into smaller subsets while incrementally developing the associated tree.



K-Nearest Neighbors

A simple algorithm that stores all available cases and predict the numerical target based on a similarity measure.

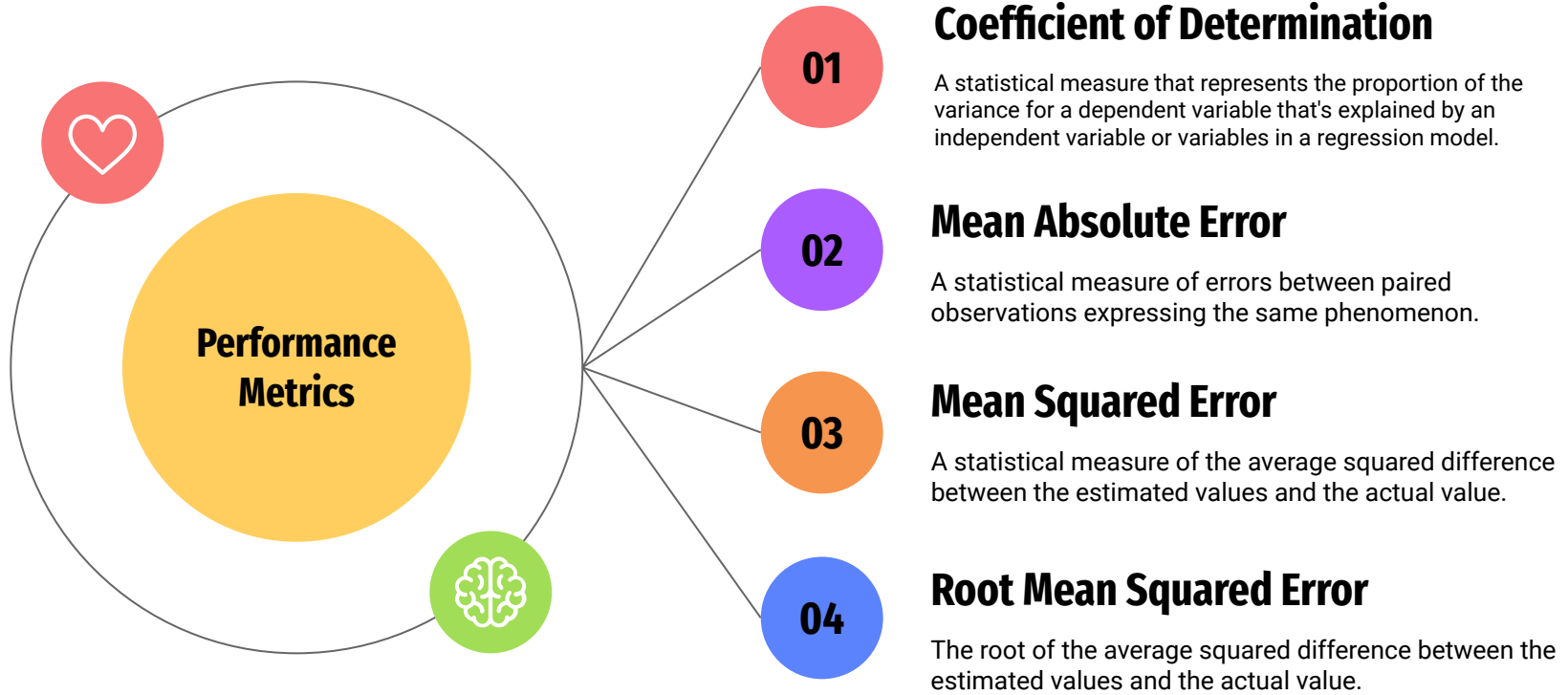
Support Vector Regressor

A supervised learning algorithm that is used to predict discrete values

Gradient Boosting

An ML technique which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.

Experimental Study



Results & Analysis

Anger							Joy						
	Linear Regression	Random Forest	Decision Tree	KNN	SVR	Gradient Boost		Linear Regression	Random Forest	Decision Tree	KNN	SVR	Gradient Boost
R2	-7.3828e+21	-0.0788643	-0.182579	-0.140585	-0.0153836	-0.0223897	R2	-1.96711e+22	-0.0910428	-0.159834	-0.384159	-0.0149771	-0.0109779
MAE	3.48474e+09	0.170682	0.177707	0.17367	0.165721	0.16674	MAE	3.18182e+09	0.162775	0.167744	0.182052	0.161567	0.159749
MSE	2.92349e+20	0.0427216	0.0468286	0.0451657	0.0402079	0.0404853	MSE	7.22398e+20	0.0400672	0.0425935	0.0508316	0.0372738	0.0371269
RMSE	1.70982e+10	0.206692	0.216399	0.212522	0.200519	0.20121	RMSE	2.68775e+10	0.200168	0.206382	0.225459	0.193064	0.192684
Sadness							Fear						
	Linear Regression	Random Forest	Decision Tree	KNN	SVR	Gradient Boost		Linear Regression	Random Forest	Decision Tree	KNN	SVR	Gradient Boost
R2	-6.71138e+21	-0.106157	-0.207057	-0.236501	-0.00589185	-0.0225636	R2	-5.0432e+21	-0.161456	-0.227915	-0.209392	-0.0226743	-0.0947355
MAE	3.14869e+09	0.153242	0.159843	0.161546	0.146084	0.147382	MAE	2.50807e+09	0.151374	0.155281	0.156681	0.144375	0.147129
MSE	2.18221e+20	0.0359667	0.0392475	0.0402049	0.0327066	0.0332487	MSE	1.58992e+20	0.036616	0.0387112	0.0381272	0.0322408	0.0345126
RMSE	1.47723e+10	0.189649	0.19811	0.200511	0.18085	0.182342	RMSE	1.26092e+10	0.191353	0.196752	0.195262	0.179557	0.185776

Exploratory Data Analysis: Word Clouds

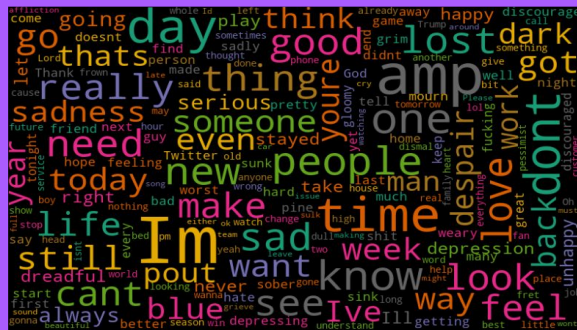
01 Joy



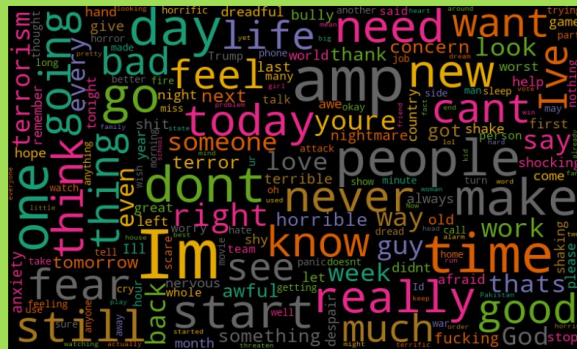
02 Anger



03 Sadness

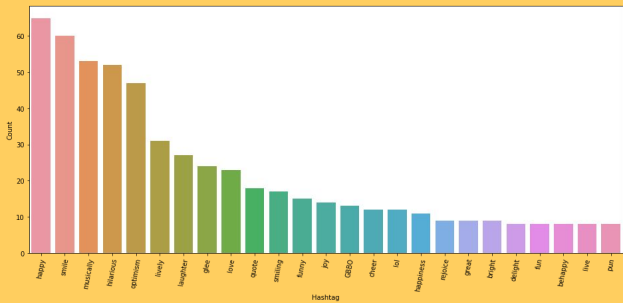


04 Fear

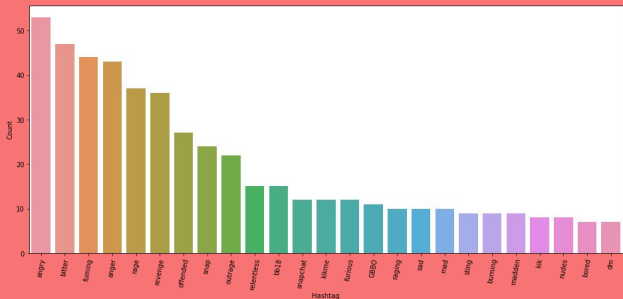


Exploratory Data Analysis: Bar Plots

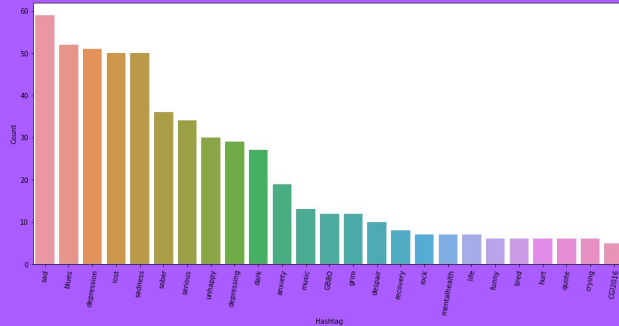
01 Joy



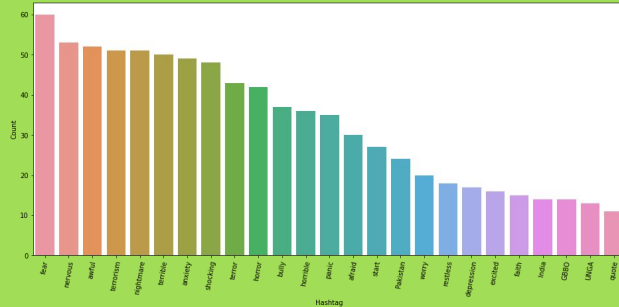
02 Anger



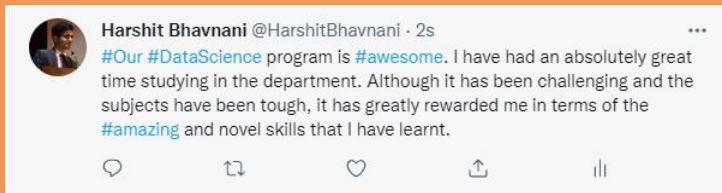
03 Sadness



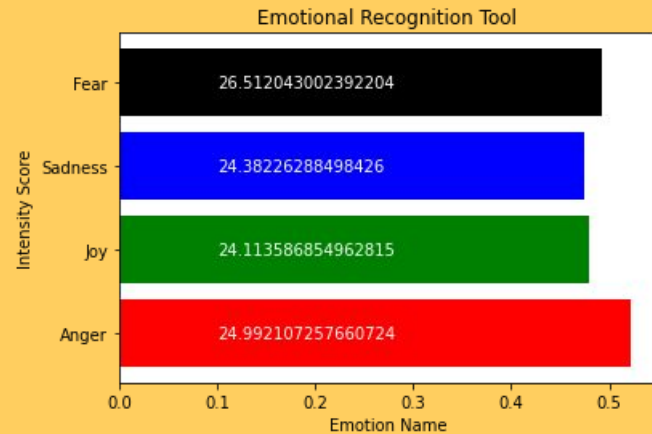
04 Fear



Application



Input



Output