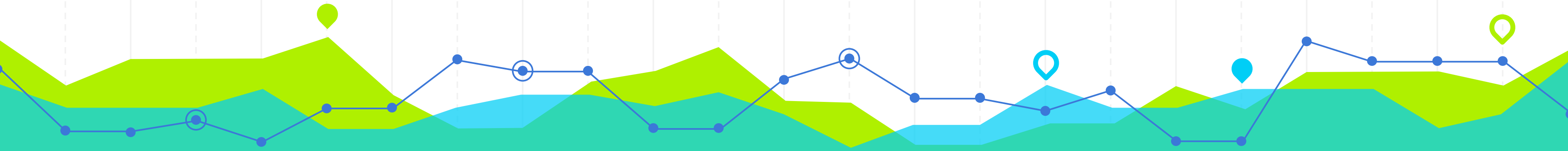
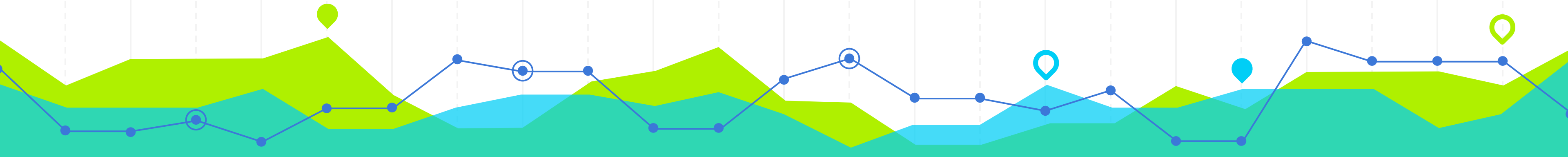


GREENPASS Sentiment Analysis

Cancello Tortora Giuseppe
Macrì Armando

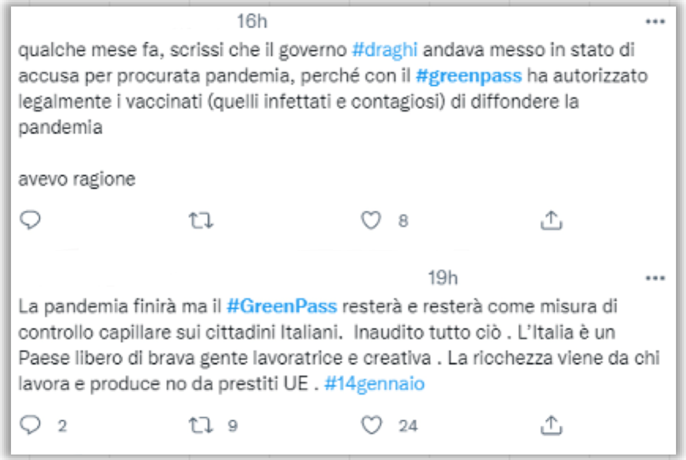
a.a. 2021-2022



INTRODUCTION 1

Goal of the project

The goal of this project is to develop a monitoring analysis in order to extract useful information and retrieve what is the popular opinion about **Green Pass**.



16h

qualche mese fa, scrissi che il governo #draghi andava messo in stato di accusa per procurata pandemia, perché con il #greenpass ha autorizzato legalmente i vaccinati (quelli infettati e contagiosi) di diffondere la pandemia

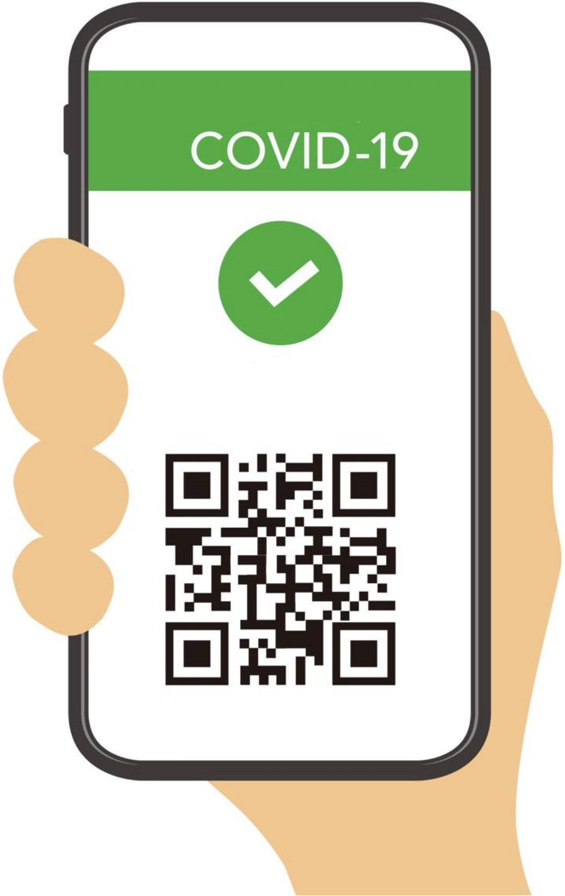
avevo ragione

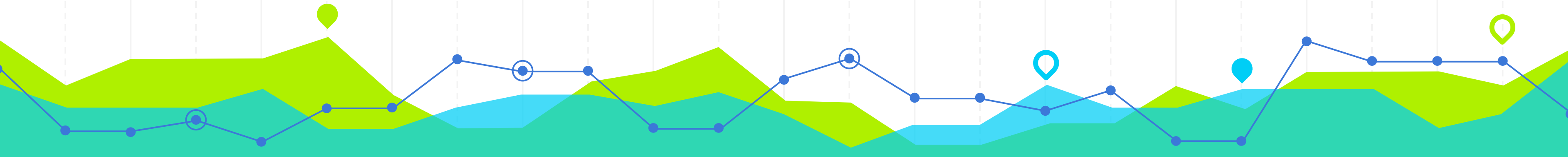
8

19h

La pandemia finirà ma il #GreenPass resterà e resterà come misura di controllo capillare sui cittadini Italiani. Inaudito tutto ciò . L'Italia è un Paese libero di brava gente lavoratrice e creativa . La ricchezza viene da chi lavora e produce no da prestiti UE . #14gennaio

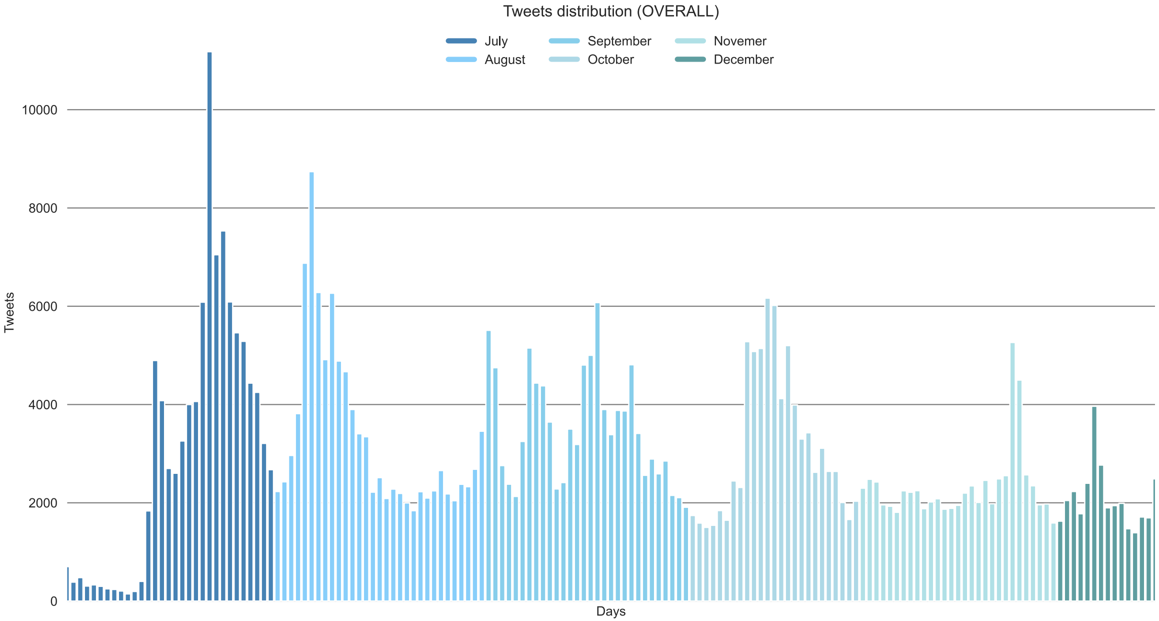
2 9 24





DATASET 2

DATASET - BUILDING



- Tweets are collected by using a Python script

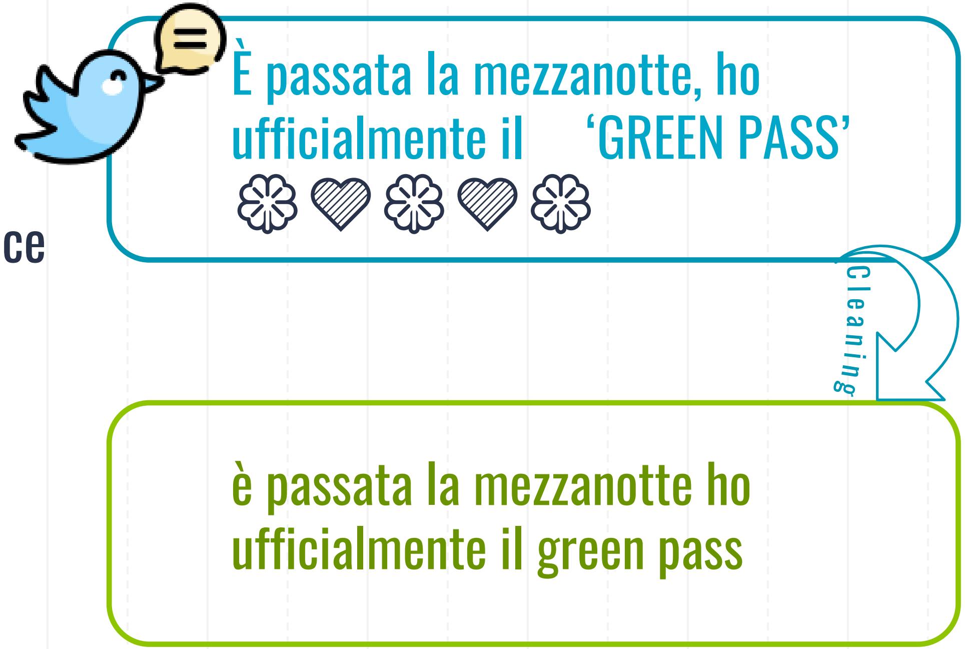
- From July 1st to December 15th

- ~ 500.000+ tweets

DATASET - CLEANING

- Keep only italian tweets
- Removing URL, mentions, emoticons
- Replacing multiple spaces with single space
- Removing punctuaction marks
- Lower case

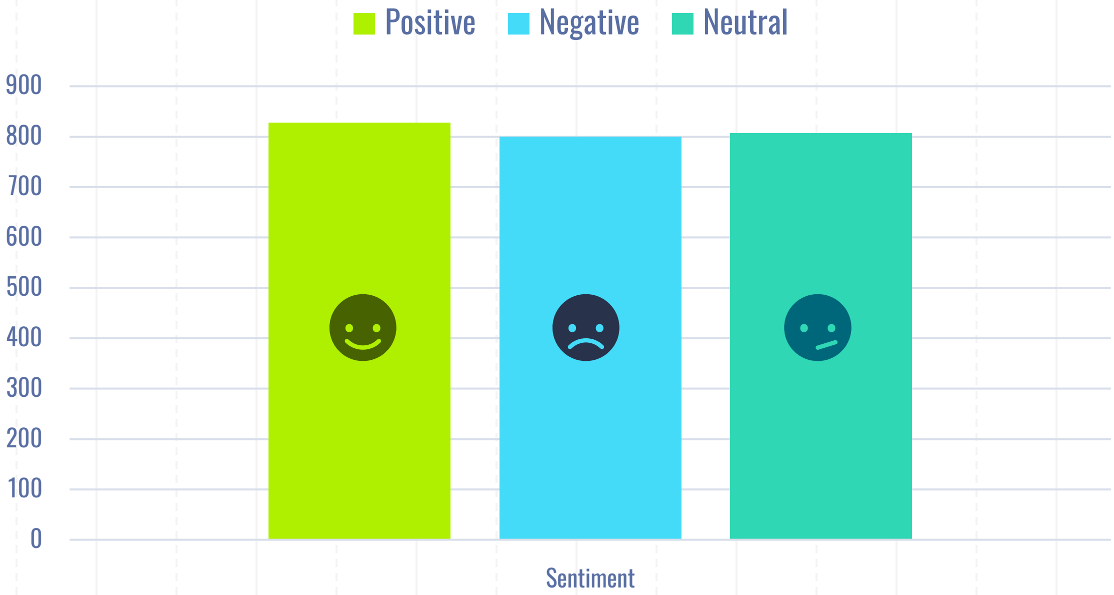
Clean tweets content from useless stuff and try to **standardize** them as much as possible.



È passata la mezzanotte, ho ufficialmente il   ‘GREEN PASS’ ❀♡❀♡❀

è passata la mezzanotte ho ufficialmente il green pass

DATASET – TRAINING SET



- ~ 2400 tweets labelled by hand
- Balanced dataset, about 800 instances for each class
- Timeline for training set: 01/07/2021 to 22/07/2021

DATA PREPOCESSING

TOKENIZATION

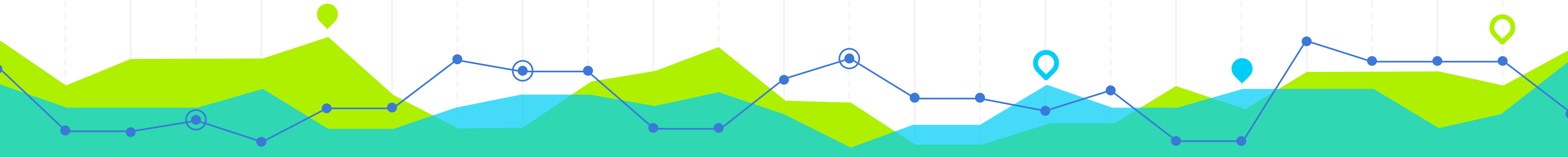
STOP-WORDS FILTERING

Words from file (stopwords_it)

STEMMING

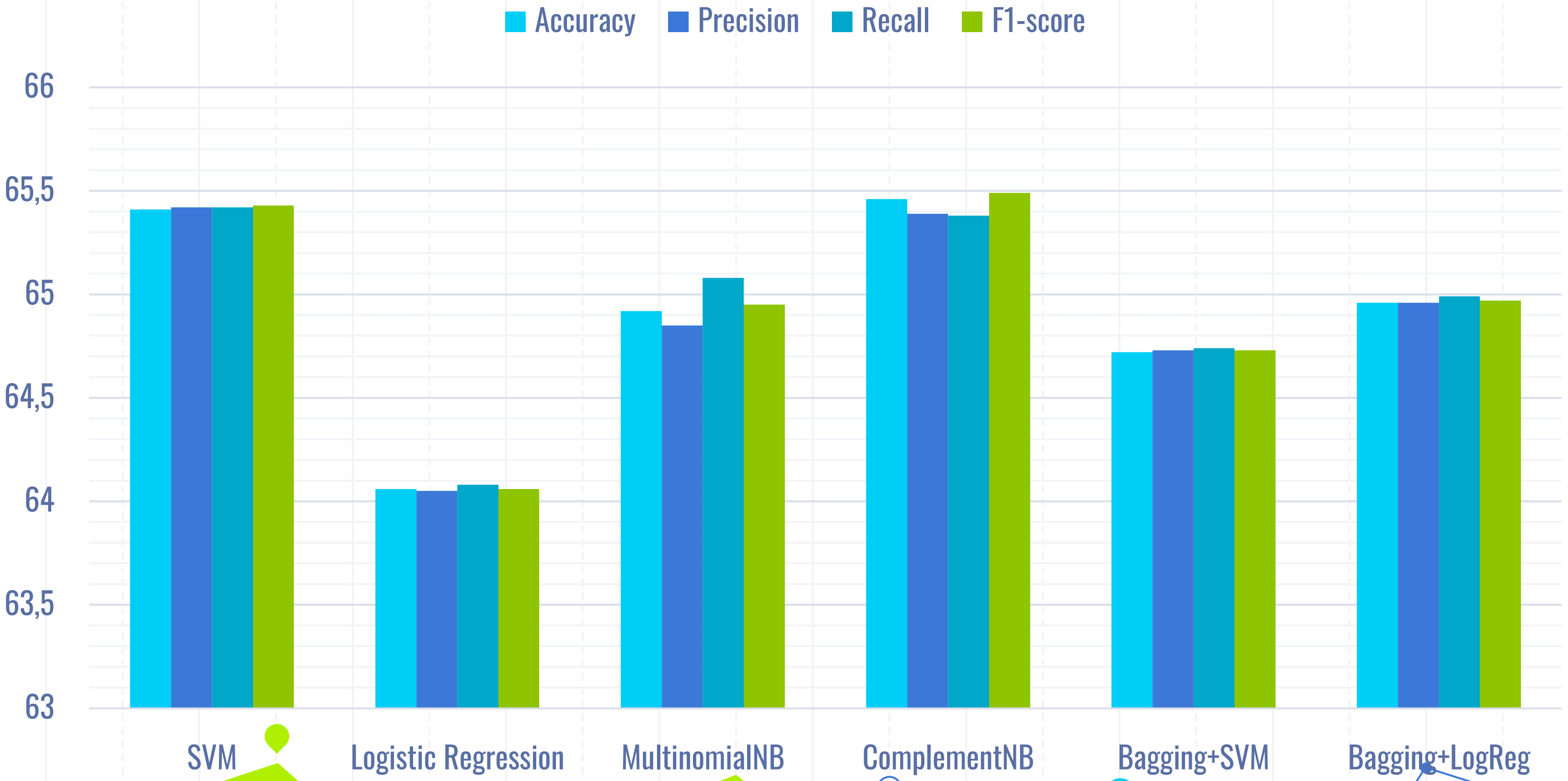
FEATURES EXTRACTION

TFIDF



CLASSIFICATION 3

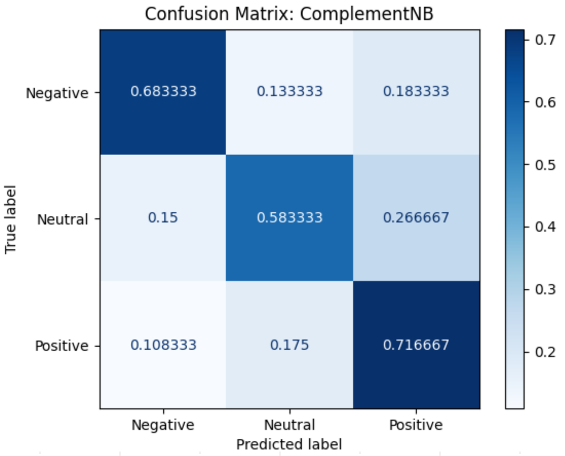
MODEL PERFORMANCE ON TRAINING SET



MODEL SELCTION USING PAIRED T-TEST

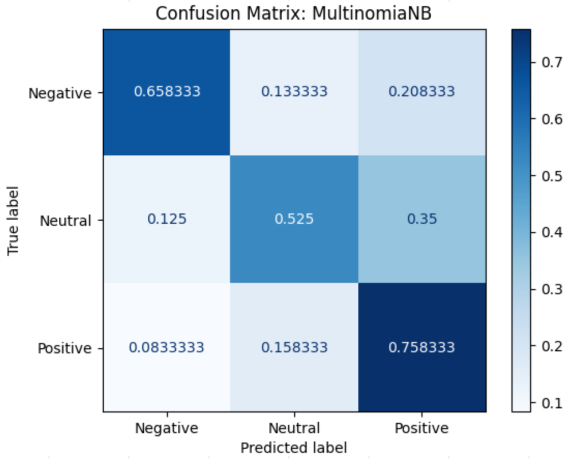
CLASSIFIER	LOG REG	MULTI_NB	COMP_NB	BAG_SVM	BAG_LOG
SVM	4.3883 0.6467 0.6522	1.5077 0.6494 0.6522	-1.2245 0.6545 0.6522	7.6961 0.6456 0.6522	7.7386 0.6444 0.6522
LOG REG		-1.5650 0.6494 0.6467	-6.0565 0.6545 0.6467	0.8210 0.6456 0.6467	2.6656 0.6444 0.6467
MULTU_NB			-3.459 0.6545 0.6494	1.7161 0.6456 0.6494	2.2671 0.6444 0.6494
COMP_NB				4.3025 0.6456 0.6545	5.5733 0.6444 0.6545
BAG_SVM					1.0879 0.6444 0.6456

CLASSIFICATION – SELECTION



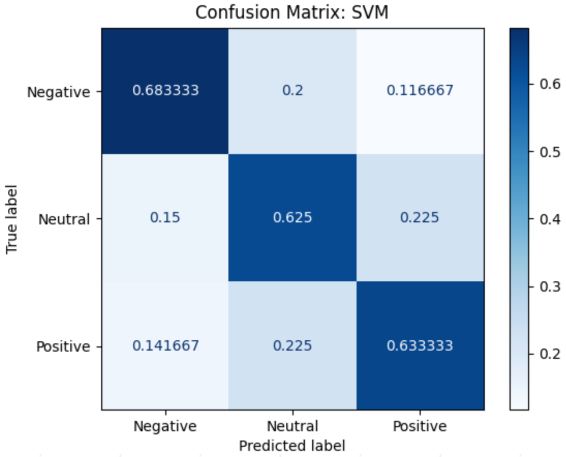
Confusion Matrix: ComplementNB

True label \ Predicted label	Negative	Neutral	Positive
Negative	0.683333	0.133333	0.183333
Neutral	0.15	0.583333	0.266667
Positive	0.108333	0.175	0.716667



Confusion Matrix: MultinomiaNB

True label \ Predicted label	Negative	Neutral	Positive
Negative	0.658333	0.133333	0.208333
Neutral	0.125	0.525	0.35
Positive	0.0833333	0.158333	0.758333



Confusion Matrix: SVM

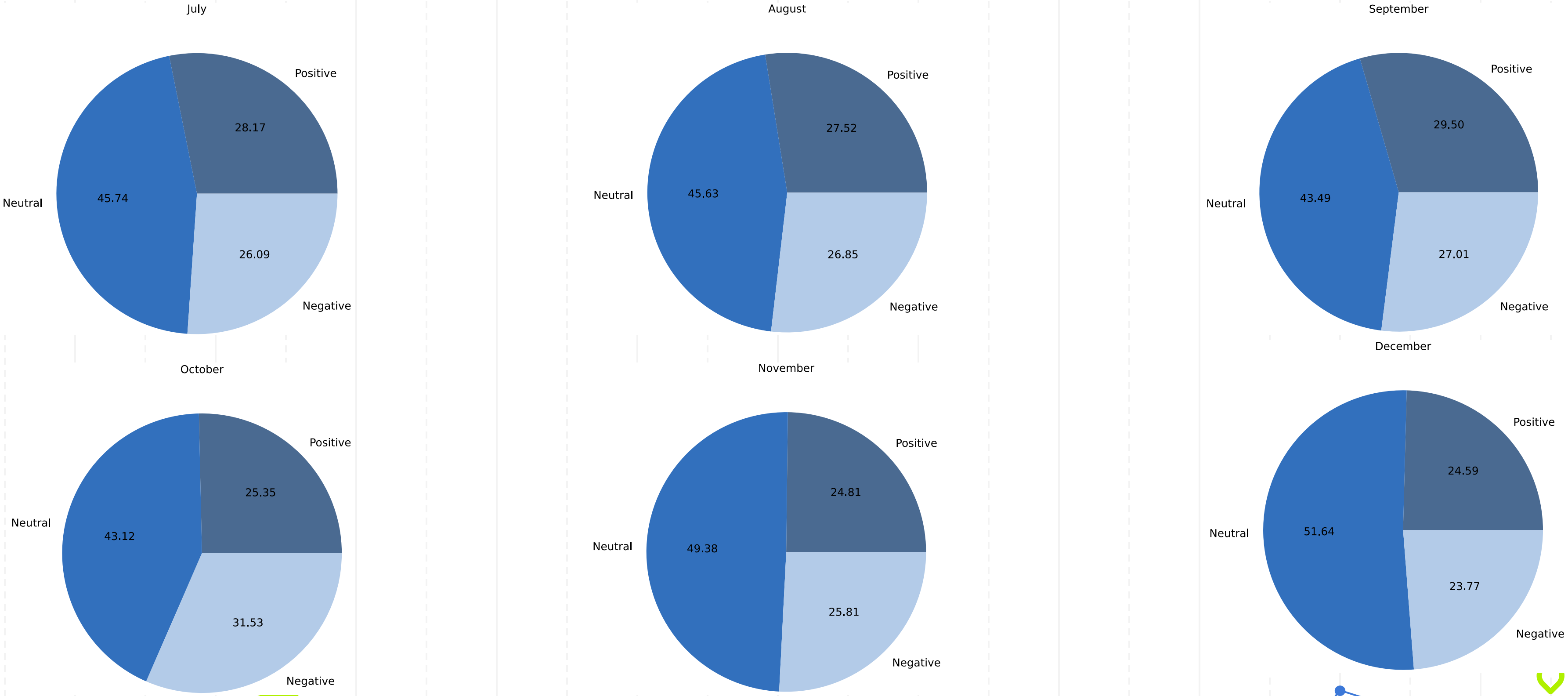
True label \ Predicted label	Negative	Neutral	Positive
Negative	0.683333	0.2	0.116667
Neutral	0.15	0.625	0.225
Positive	0.141667	0.225	0.633333

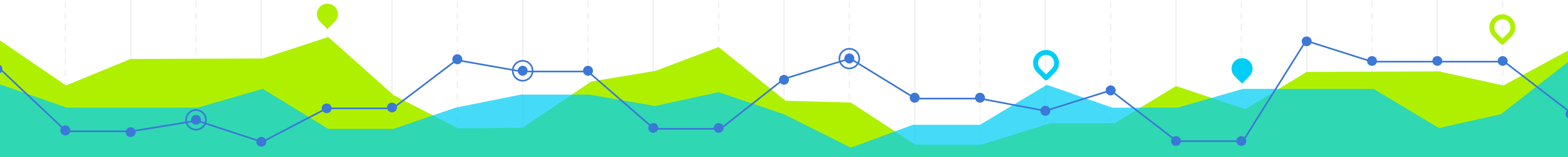
1 66,11%

2 64,72%

3 64,71%

DISTRIBUTION OF SENTIMENT

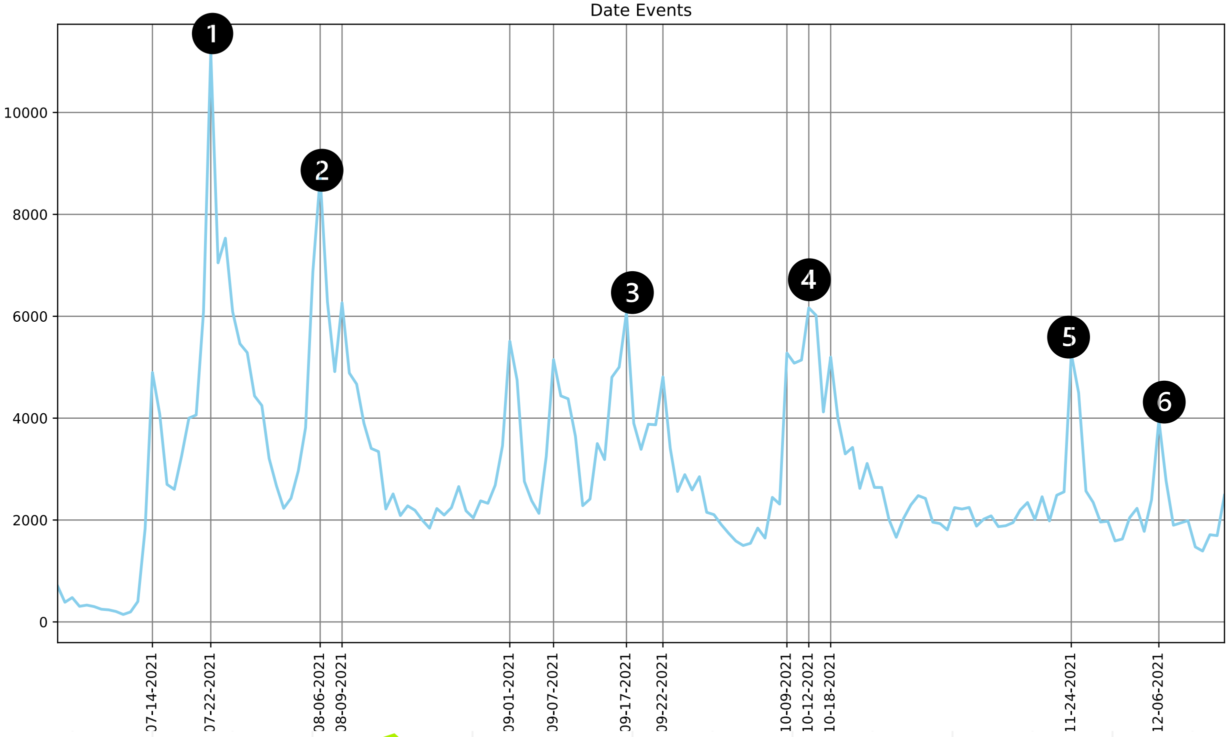




MONITORING ANALYSIS

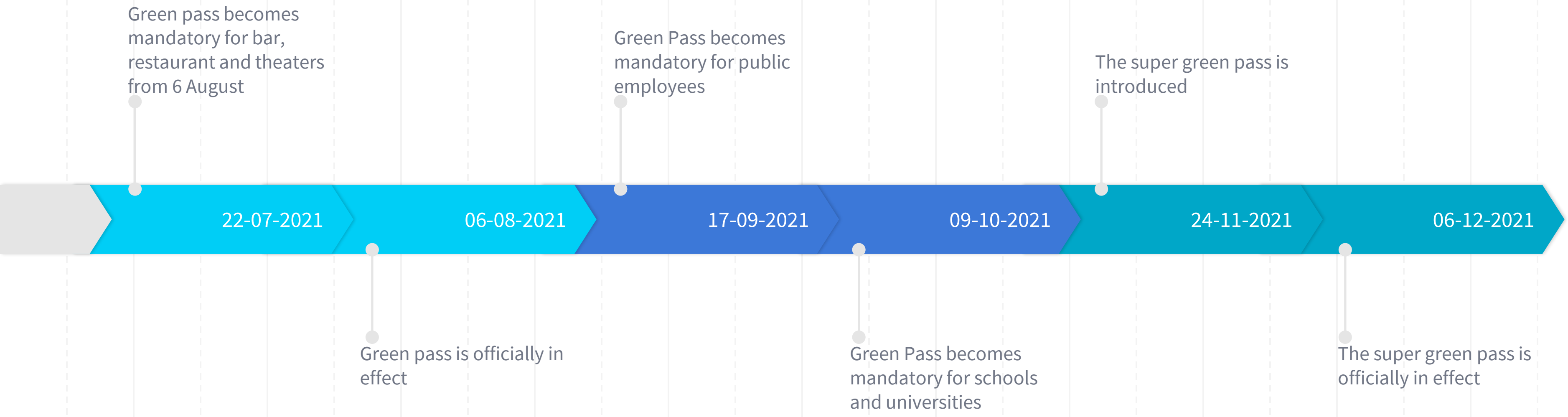
4

MONITORING ANALYSIS – TIMELINE



- Found 13 key events on the timeline considered
- Peak of tweets of those events
- We focused on 6 events

EVENTS



22-07-2021

Green pass becomes mandatory for bar, restaurant and theaters from 6 August

06-08-2021

Green pass is officially in effect

17-09-2021

Green Pass becomes mandatory for public employees

09-10-2021

Green Pass becomes mandatory for schools and universities

24-11-2021

The super green pass is introduced

06-12-2021

The super green pass is officially in effect

MONITORING ANALYSIS – SCHEMAS

3 different learning settings!

INCREMENTAL MODEL

Trained with the initial training set and all the hand labelled data of all the previous events before testing on a new event.

SLIDING MODEL

Retrained each time with the most recent 2400 tweets, removing the oldest 360 and adding the newest 360.

STATIC MODEL

The initial training set composed by 2400 tweets

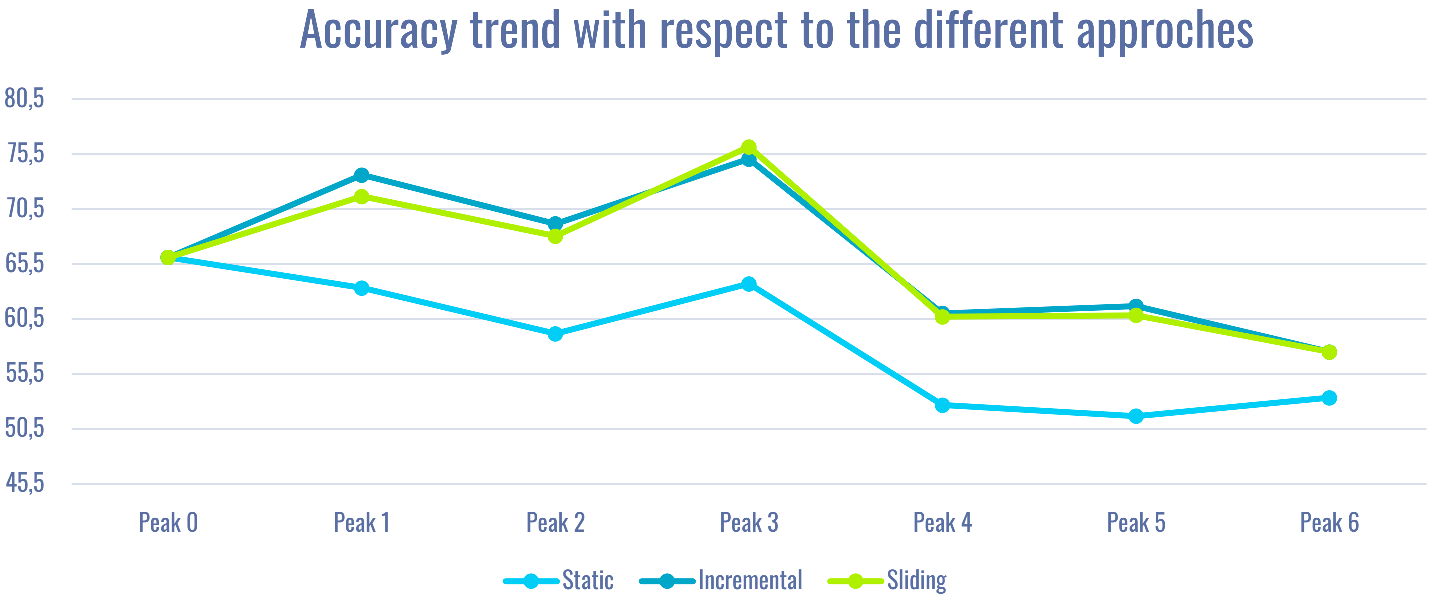
MONITORING ANALYSIS – RESULTS

SHEMAS	Accuracy (%)		
	Incremental	Sliding	Static
Peak 1	73,61	71,66	63,33
Peak 2	69,16	68,05	59,16
Peak 3	75,06	76,17	63,71
Peak 4	61,01	60,71	52,67
Peak 5	61,66	60,83	51,66
Peak 6	57,50	57,50	53,33
Average	**66,33**	**65,82**	**57,31**

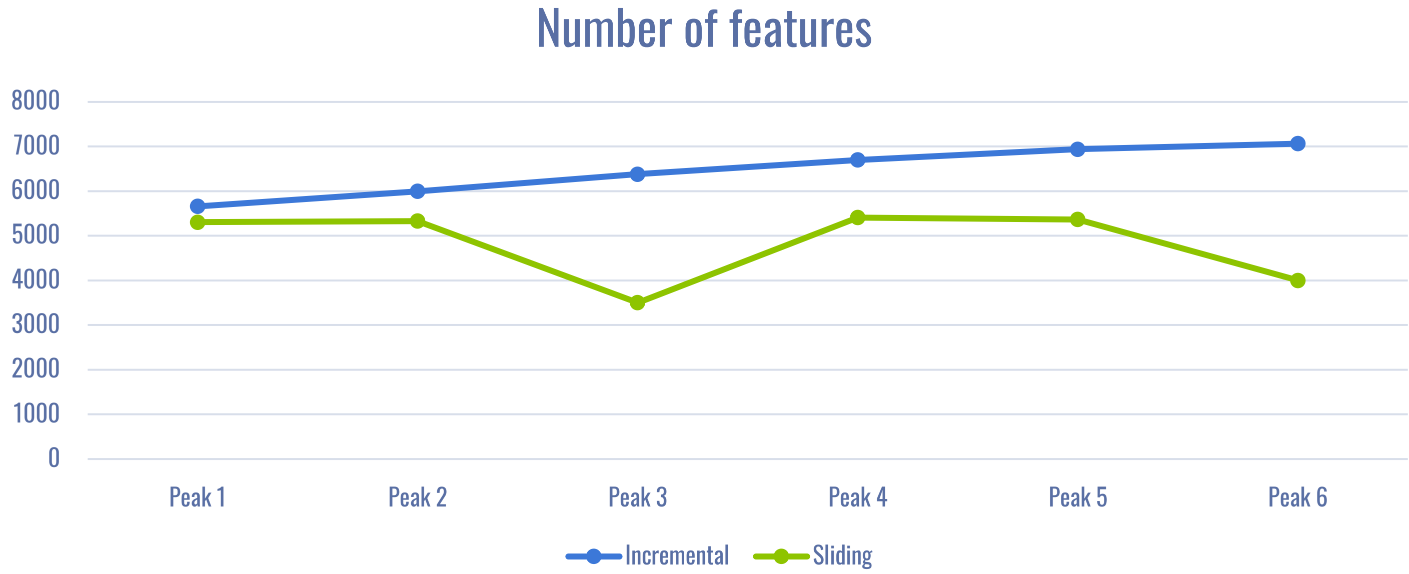
These are the results in terms of accuracy of the three considered approaches.

Incremental and sliding model perform quite well, while the static model is affected by the concept drift.

MONITORING ANALYSIS - RESULTS

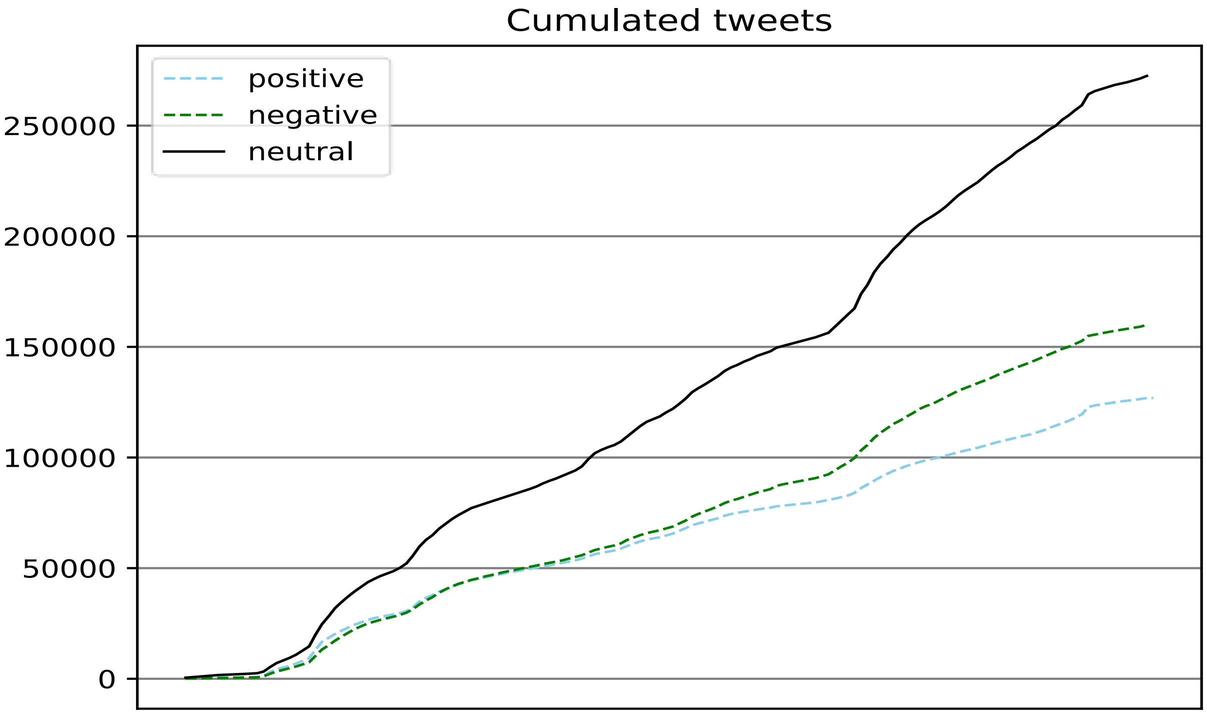


MONITORING ANALYSIS - RESULTS

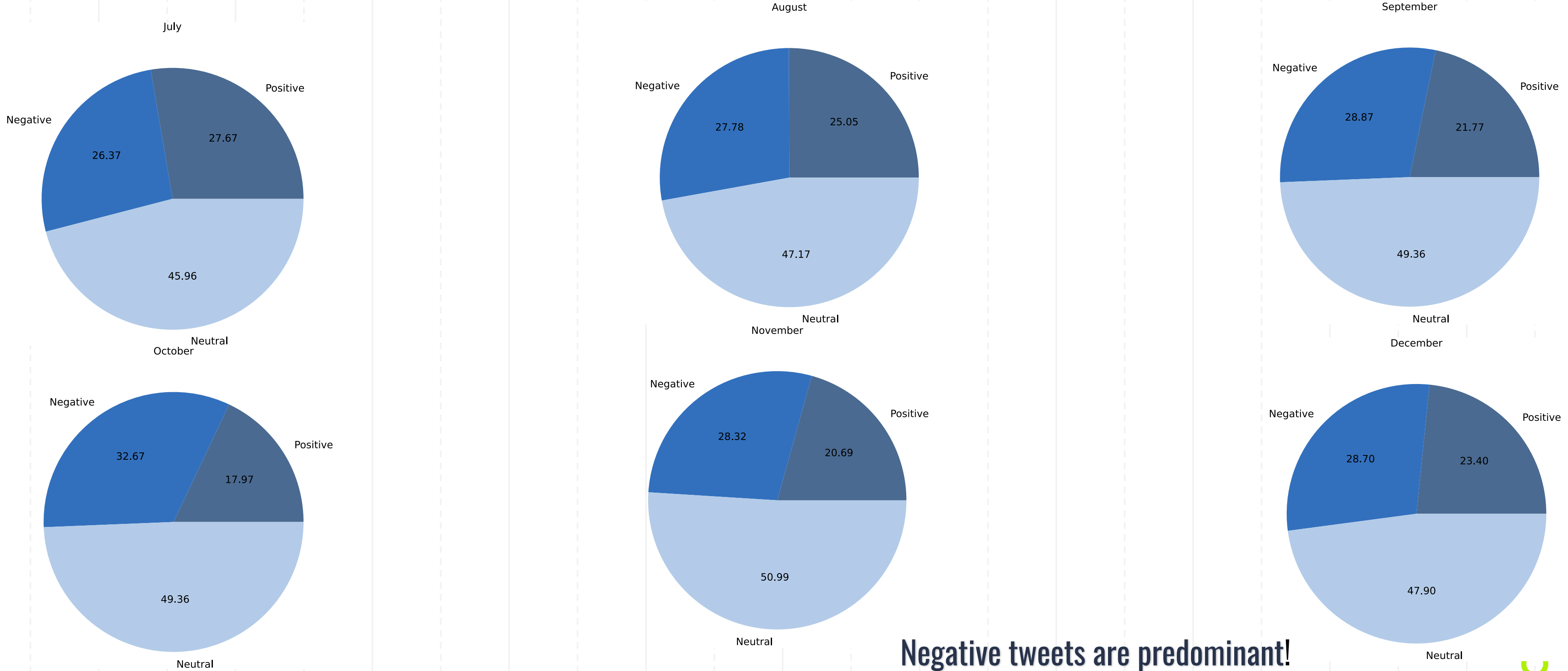


We select the *sliding* approach to conclude our analysis because, on average, it gave us results comparable to those obtained with the incremental approach, and it requires, also, a smaller number of features.

MONITORING ANALYSIS - CONCLUSION



DISTRIBUTION OF SENTIMENT



Negative tweets are predominant!

THANKS!