

# SENTIMENT ANALYSIS FILM **SPIDERMAN NO WAY HOME** PADA **TWITTER**

Adhitya Primandhika  
19/444033/TK/49229





# Background

Latar belakang saya tertarik untuk melakukan analisis sentimen masyarakat pada platform **Twitter** terhadap film **Spiderman No Way Home** adalah karena film ini merupakan salah satu film yang ditunggu-tunggu oleh pecinta film superhero khususnya fans Marvel (termasuk saya).

Film ini ditunggu-tunggu karena dikabarkan Spiderman versi **Tom Holland** selain bertemu dengan villain dari Spiderman versi sebelumnya akan bertemu pula dengan **Spiderman** versi pendahulunya yaitu versi **Tobey Maguire** dan versi **Andrew Garfield**.

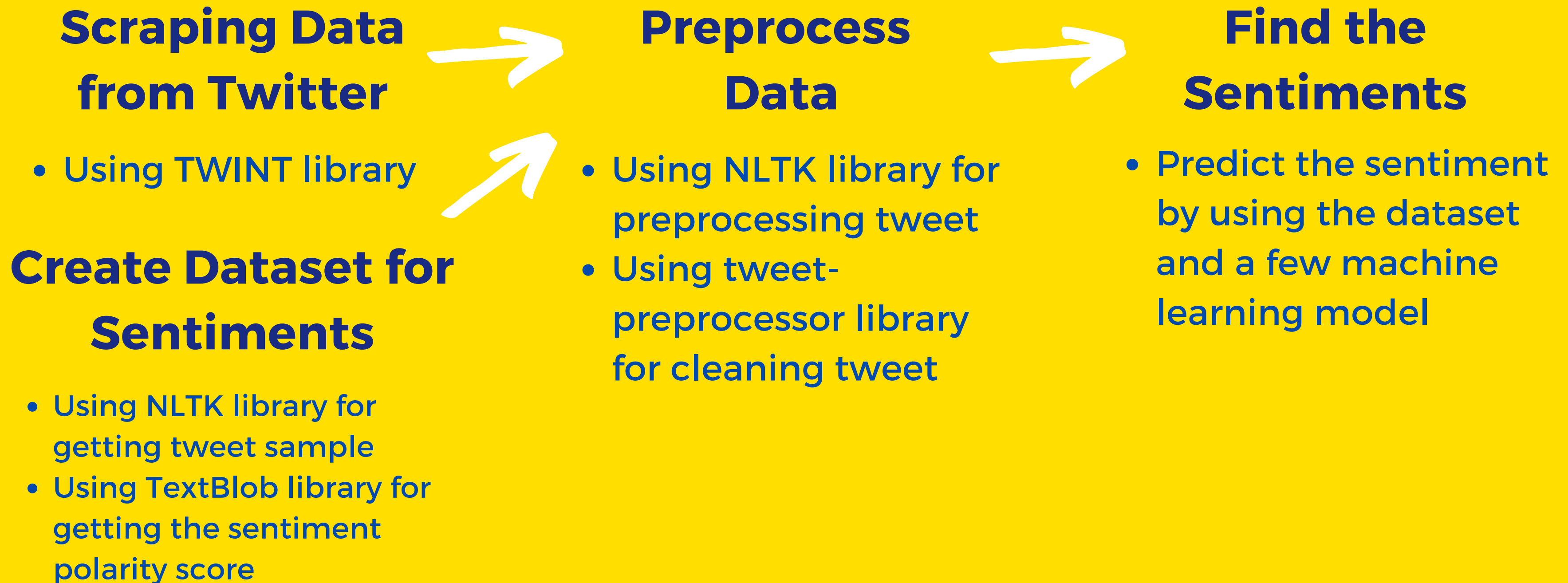


# Purpose

Tujuan dari dilakukannya analisis sentimen pada **Twitter** terkait film **Spiderman No Way Home** melihat seperti apa **sentimen yang ada pada masyarakat** khususnya untuk pengguna Twitter terkait film **Spiderman No Way Home** yang akan dirilis di Indonesia pada 15 Desember 2021 mendatang dan pada 17 Desember 2021 untuk seluruh dunia.



# Process / Schema for The Sentiment Analysis





# Functions

## Preprocess & Cleaning

```
def preprocess(text, stopwords):  
    if not stopwords:  
        stopwords = []  
    text = text.lower() # Make the text lowercase  
    text = p.clean(text) # Clean text from any mention, emoji, hashtag, reserve words(such as FAV, RT), smiley, and  
url  
    text = text.strip().replace("\n", "")  
    text = re.sub("[^a-zA-Z0-9]", " ", text)  
    tokens = [w for w in text.split() if not w in stopwords]  
    text = " ".join(tokens)  
    return text
```

## Get Sentiment

```
def get_sentiment(text):  
    # Classify sentiment based on sentiment polarity score  
    sentiment = ""  
    if TextBlob(text).sentiment.polarity > 0:  
        sentiment = "positive"  
    elif TextBlob(text).sentiment.polarity < 0:  
        sentiment = "negative"  
    else:  
        sentiment = "neutral"  
    return sentiment
```

# Functions

## Evaluate Model

```
def model_evaluate(model):  
    # Predict values for Test dataset  
    y_pred = model.predict(X_test)  
  
    # Print the evaluation metrics for the dataset  
    print(classification_report(y_test, y_pred))
```

## Predict The Sentiment

```
def predict_sentiments(vectoriser, model, texts):  
    # Predict the sentiment  
    for item in texts:  
        item = item.lower() # Make the text lowercase  
        item = preprocess(item, stopwords_list) # Preprocess the text  
    textdata = vectoriser.transform(texts) # Vectorise the text with vectoriser  
    sentiments = model.predict(textdata) # Predict the sentiment from the text  
    scores = cross_val_score(SVCmodel, textdata, sentiments, cv=5) # Do cross validation  
    print("Cross Validation Score : {}".format(scores))  
    df = pd.DataFrame(texts, columns=["tweet"])  
    df["sentiment"] = sentiments  
    return df
```

# Scraping Data



```
c = twint.Config()
c.Store_json = True
c.Output = f"{PROJECT_NAME}.json"
c.Search = "#SpiderManNoWayHome"
c.Until = "2021-12-10"
c.Lang = "en"
c.Resume = "resume.txt"

c.Hide_output = True
```

Untuk **scraping data** yang saya lakukan pada platform **Twitter**, saya melakukan **scraping data** dengan melakukan konfigurasi untuk mencari semua tweet yang menggunakan hashtag **#SpiderManNoWayHome** dengan batas tanggalnya adalah sampai tanggal **10 Desember 2021** tepat seminggu sebelum perilisan film tersebut. Selain itu, saya juga hanya akan mengambil **tweet** yang **berbahasa inggris** saja.



# Tweets Data from Scraping Twitter

Setelah kita melakukan **scraping data** pada **Twitter** dengan konfigurasi yang sudah kita tentukan, kita akan mendapatkan data berupa **tweet-tweet** yang menggunakan hashtag **#SpiderManNoWayHome**. Untuk data yang didapatkan merupakan tweet dari tanggal **2 Desember 2021** sampai tanggal **10 Desember 2021**. Jumlah tweet yang didapatkan adalah sebanyak **20.755 tweet**. Berikut merupakan cuplikan data tweet yang didapatkan.

	id	created_at	date	tweet	language
0	1469094336952475651	2021-12-10 06:59:19+07:00	2021-12-09	Can't believe we're about to witness an Avengers Endgame level movie NEXT WEEK #SpiderManNoWayHome its been a while tbh..!	en
1	1469093770918572039	2021-12-10 06:57:04+07:00	2021-12-09	Im waiting 🕷️ #SpiderManNoWayHome @MarvelStudios <a href="https://t.co/AnFH8pET5w">https://t.co/AnFH8pET5w</a>	en
2	1469093536628985857	2021-12-10 06:56:08+07:00	2021-12-09	One week away #SpiderManNoWayHome <a href="https://t.co/eVIQI9wvNm">https://t.co/eVIQI9wvNm</a>	en
3	1469093126367158272	2021-12-10 06:54:30+07:00	2021-12-09	Show me who you would cast as Harry Osborn in the MCU after #SpiderManNoWayHome	en
4	1469092855377539075	2021-12-10 06:53:25+07:00	2021-12-09	@ComicBook New Scene from the #DoctorOctopus streetfight for #SpiderManNoWayHome we see here a British Spiderman , Sparking his marbles on the street , something you never saw in original Spiderman Books , cause that wouldnt work with anything spiderman does , " Nuts-on-Fire "	en
...	...	...	...	...	...
20750	1466130115734253572	2021-12-02 02:40:33+07:00	2021-12-01	Guess what month it is 🍷 #SpiderMan #SpiderManNoWayHome <a href="https://t.co/HbbOm22nMA">https://t.co/HbbOm22nMA</a>	en
20751	1466130099426803712	2021-12-02 02:40:29+07:00	2021-12-01	15 Days To Go 📅 #SpiderManNoWayHome <a href="https://t.co/9OPkLG3b9N">https://t.co/9OPkLG3b9N</a>	en
20752	1466129865992814603	2021-12-02 02:39:34+07:00	2021-12-01	@SonyPictures We know you could do better. 🕷️ #SpiderManNoWayHome <a href="https://t.co/qqCsJq5BKC">https://t.co/qqCsJq5BKC</a>	en
20753	1466129479420497930	2021-12-02 02:38:01+07:00	2021-12-01	Calling all you Spidey fans, check out our exclusive #SpiderManNoWayHome combo - available to pre-order online during ticket purchase or in cinema from 15 December. 🕷️ <a href="https://t.co/Fn2IL3452b">https://t.co/Fn2IL3452b</a> <a href="https://t.co/XGAI7jeAeA">https://t.co/XGAI7jeAeA</a>	en
20754	1466129415356784644	2021-12-02 02:37:46+07:00	2021-12-01	@_Nick_Nation_ @WHOTheBckstreet @nickcarter @Cure4thekids Thank you for the update! I wasn't sure if I was going to be able to participate in Nick's #Home4TheHolidays event since I had already secured coveted, premier-night, #SpiderManNoWayHome tickets at my local #IMAX happening that same night! But now I can enjoy BOTH!!! 🍷 <a href="https://t.co/iAjfmSiKFZ">https://t.co/iAjfmSiKFZ</a>	en
20755 rows × 5 columns					



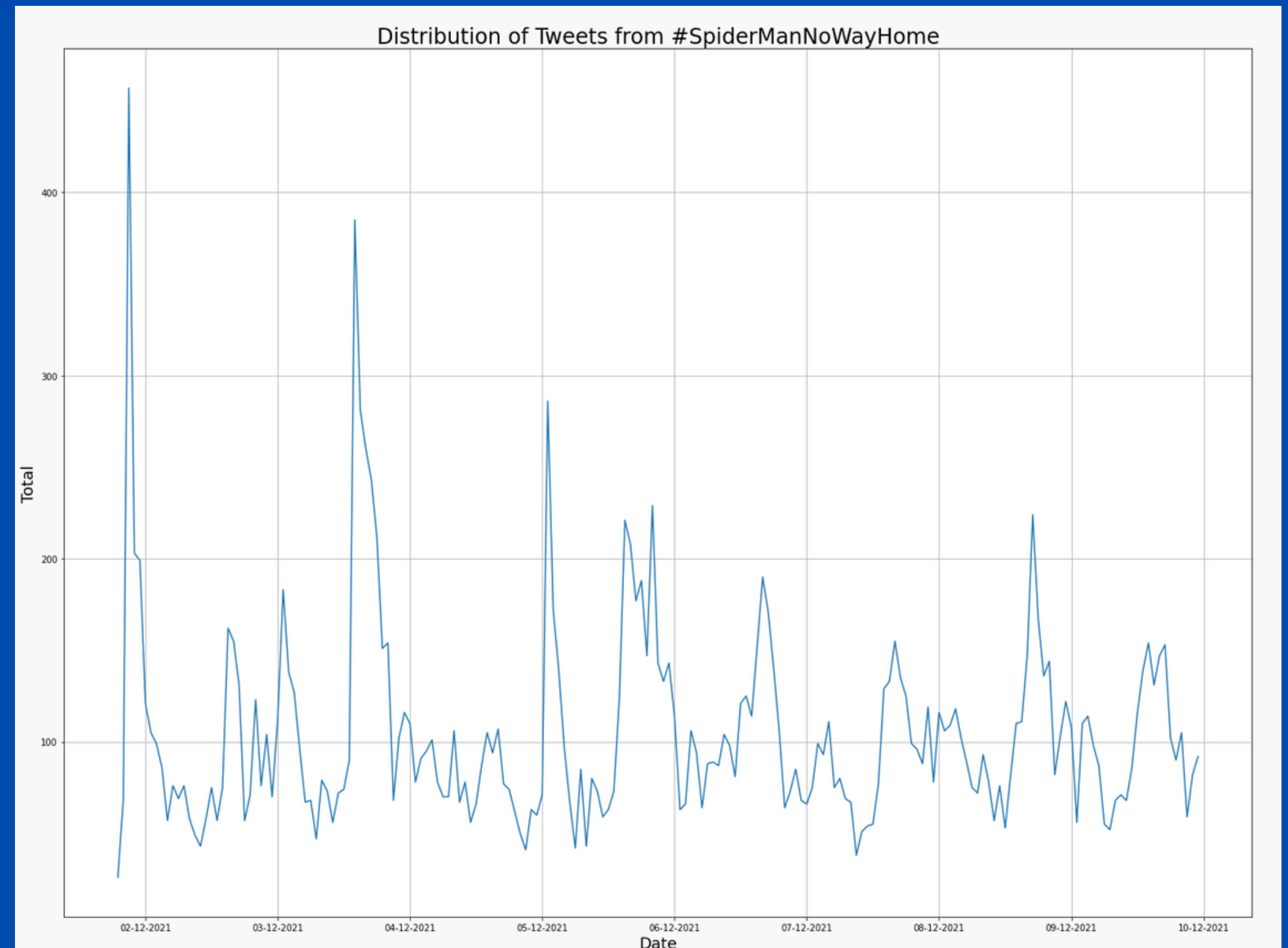
# Tweets Data from Scraping Twitter

Setelah kita dapatkan data **tweet** tersebut, barulah kita lakukan **preprocessing** pada data-data tersebut agar kita dapat mencari sentimennya. Berikut cuplikan data yang sudah dilakukan **preprocessing** dengan fungsi **preprocess()** yang telah dibuat di awal.

	id	created_at	date	tweet	language
0	1469094336952475651	2021-12-10 06:59:19+07:00	2021-12-09	Cant believe were about to witness an Avengers Endgame level movie NEXT WEEK its been a while tbh	en
1	1469093770918572039	2021-12-10 06:57:04+07:00	2021-12-09	Im waiting	en
2	1469093536628985857	2021-12-10 06:56:08+07:00	2021-12-09	One week away	en
3	1469093126367158272	2021-12-10 06:54:30+07:00	2021-12-09	Show me who you would cast as Harry Osborn in the MCU after	en
4	1469092855377539075	2021-12-10 06:53:25+07:00	2021-12-09	New Scene from the streetfight for we see here a British Spiderman Sparking his marbles on the street something you never saw in original Spiderman Books cause that wouldnt work with anything spiderman does Nuts on Fire	en
...	...	...	...	...	...
20750	1466130115734253572	2021-12-02 02:40:33+07:00	2021-12-01	Guess what month it is	en
20751	1466130099426803712	2021-12-02 02:40:29+07:00	2021-12-01	15 Days To Go	en
20752	1466129865992814603	2021-12-02 02:39:34+07:00	2021-12-01	We know you could do better	en
20753	1466129479420497930	2021-12-02 02:38:01+07:00	2021-12-01	Calling all you Spidey fans check out our exclusive combo available to pre order online during ticket purchase or in cinema from 15 December	en
20754	1466129415356784644	2021-12-02 02:37:46+07:00	2021-12-01	Thank you for the update I wasn t sure if I was going to be able to participate in Nick s event since I had already secured coveted premier night tickets at my local happening that same night But now I can enjoy BOTH	en
20755 rows × 5 columns					

# Tweet Distribution

Dari hasil scraping **Twitter** yang telah saya lakukan, didapatkan distribusi dari tweet yang menggunakan hashtag **#SpiderManNoWayHome** adalah seperti pada grafik berikut. Terlihat pada **awal** bulan **Desember** terdapat **jumlah tweet tertinggi** yang menggunakan hashtag **#SpiderManNoWayHome**.



# Create Dataset

Setelah kita sudah melakukan **scraping data Twitter**, selanjutnya kita akan membuat sebuah **dataset** yang digunakan untuk **membuat model**. Disini akan menggunakan library NLTK untuk mendapatkan **sampel tweet positif** dan **negatif** yang nantinya akan dijadikan **1 dataframe** lalu dilakukan **preprocessing** dan dicari sentimennya berdasarkan nilai **sentiment polarity** yang didapat dengan bantuan dari library TextBlob. Berikut cuplikan sampel tweet yang sudah digabungkan dan dicari sentimennya dengan fungsi **preprocess()** dan **get\_sentiment()**.

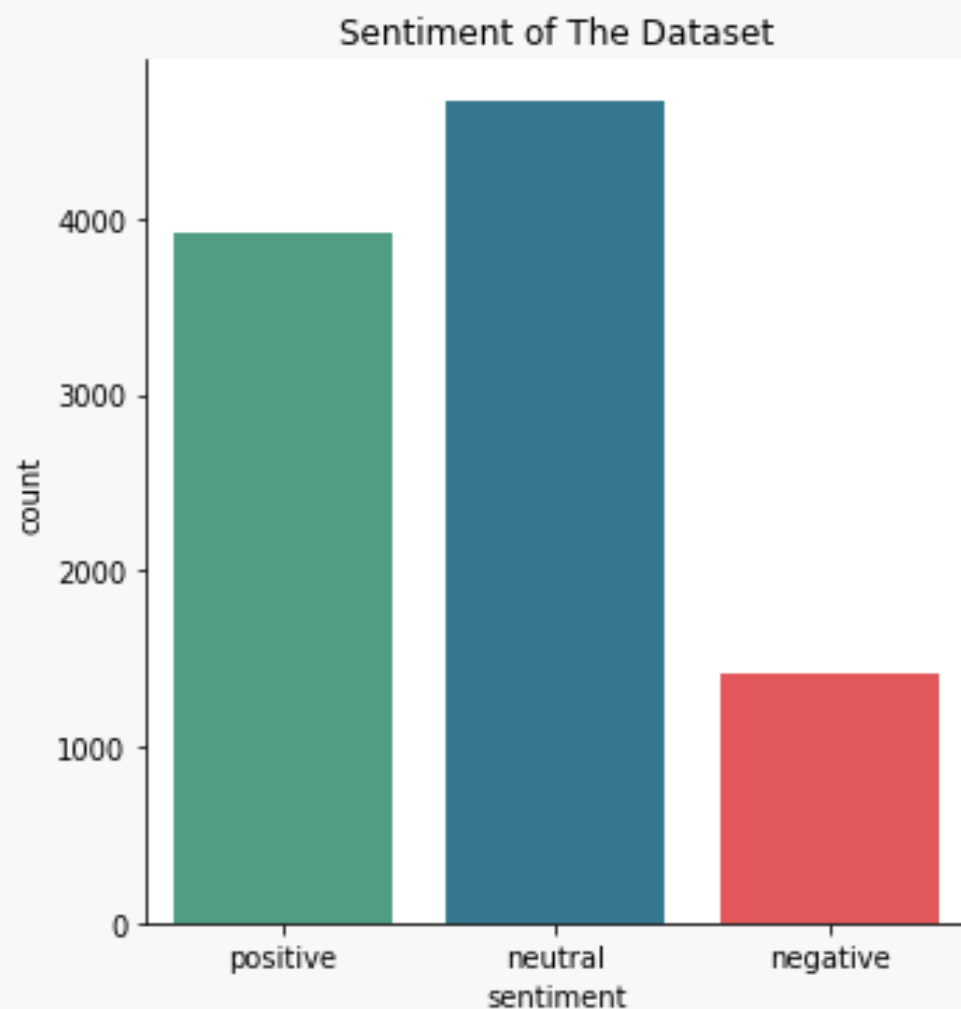
	tweet	sentiment
0	top engaged members community week	positive
1	hey james odd lease call contact centre 02392441234 able assist many thanks	positive
2	listen last night bleed amazing track scotland	positive
3	congrats	neutral
4	yeaaaah yippppy acct verified rqst succeed got blue tick mark fb profile 15 days	neutral



# Create Dataset

Selanjutnya, kita coba lihat sebaran dari sentimennya. Karena sebarannya belum merata, saya lakukan **pemerataan sampel** agar model yang dibuat akan bagus karena tidak condong ke sentimen tertentu karena adanya **oversampling**.

## Sebelum pemerataan



## Sesudah pemerataan



# Create and Choose The Best Model

Pertama kita **split dataset** yang kita punya menjadi data **train** dan data **test**

```
X_train, X_test, y_train, y_test = train_test_split(df_data["tweet"], df_data["sentiment"], test_size = 0.2, random_state = 0)
print(f'Data Split done.')
```

Selanjutnya, kita **vektorisasi** data yang kita miliki dengan teknik term-weighting **TF-IDF** dengan **normalisasi** menggunakan teknik **L2 Normalization** (norm="l2")

```
vectoriser = TfidfVectorizer(ngram_range=(1,2), max_features=500000, norm="l2")
vectoriser.fit(X_train)
print(f'Vectorizer fitted.')
print('No. of feature_words: ', len(vectoriser.get_feature_names()))
```

```
X_train = vectoriser.transform(X_train)
X_test = vectoriser.transform(X_test)
print(f'Data Transformed.')
```

# Create and Choose The Best Model

Berikut merupakan **rumus** yang digunakan pada **perhitungan** yang dilakukan oleh library Sklearn yang saya gunakan.

## Rumus TF

$$tf_{ij} = \frac{f_d(i)}{\max_{j \in d} f_d(j)}$$

## Rumus IDF

$$idf(t, D) = \log \left( \frac{N}{df(t) + 1} \right)$$

## Rumus TF-IDF

$$w_{ij} = tf_{ij} \times idf_j$$

## Rumus L2 Normalization

$$\|x\|_2 = \left( \sum_i |x_i|^2 \right)^{1/2}$$



# Create and Choose The Best Model

Selanjutnya, dengan dataset yang sudah di vektorisasi tadi kita akan melakukan **evaluasi** pada **tiap model** dengan fungsi **model\_evaluate()** yang telah dibuat di awal. Disini saya memilih 3 model, diantaranya **SVM Classifier**, **Logistic Regression**, dan **Perceptron**. Model-model tersebut umum digunakan pada analisis sentimen kecuali untuk **Perceptron**.

## SVM Classifier

	precision	recall	f1-score	support
-1	0.82	0.88	0.85	291
0	0.84	0.80	0.82	273
1	0.89	0.86	0.88	285
accuracy			0.85	849
macro avg	0.85	0.85	0.85	849
weighted avg	0.85	0.85	0.85	849

## Logistic Regression

	precision	recall	f1-score	support
-1	0.83	0.86	0.84	291
0	0.81	0.81	0.81	273
1	0.89	0.86	0.87	285
accuracy			0.84	849
macro avg	0.84	0.84	0.84	849
weighted avg	0.84	0.84	0.84	849

## Perceptron

	precision	recall	f1-score	support
-1	0.79	0.87	0.83	291
0	0.81	0.73	0.77	273
1	0.84	0.85	0.85	285
accuracy			0.82	849
macro avg	0.82	0.81	0.81	849
weighted avg	0.82	0.82	0.81	849

Dari hasil diatas, terlihat dari ketiga model yang dievaluasi ternyata **SVM Classifier** memiliki **accuracy** yang **paling tinggi** diantara ketiganya dengan nilai **85%**. Sehingga dipilihlah **SVM Classifier** ini untuk digunakan **memprediksi sentimen** pada tweet.

# Sentiment Analysis

Selanjutnya, kita akan **memprediksi sentimen** dari data yang telah kita scraping dengan menggunakan fungsi **predict\_sentiment()** yang telah dibuat di awal. Berikut kode yang digunakan.



```
result = predict_sentiments(vectoriser, SVCmodel, df_tweets["tweet"])
reverse_sentiment_mapping = {1:'positive', 0:'neutral', -1:'negative'}
result["sentiment"] = result["sentiment"].map(reverse_sentiment_mapping).astype(str)
result
```

# Classification Result (Table)

Berikut merupakan **cuplikan** dari hasil **prediksi sentimen** yang menggunakan model **SVM Classifier**.

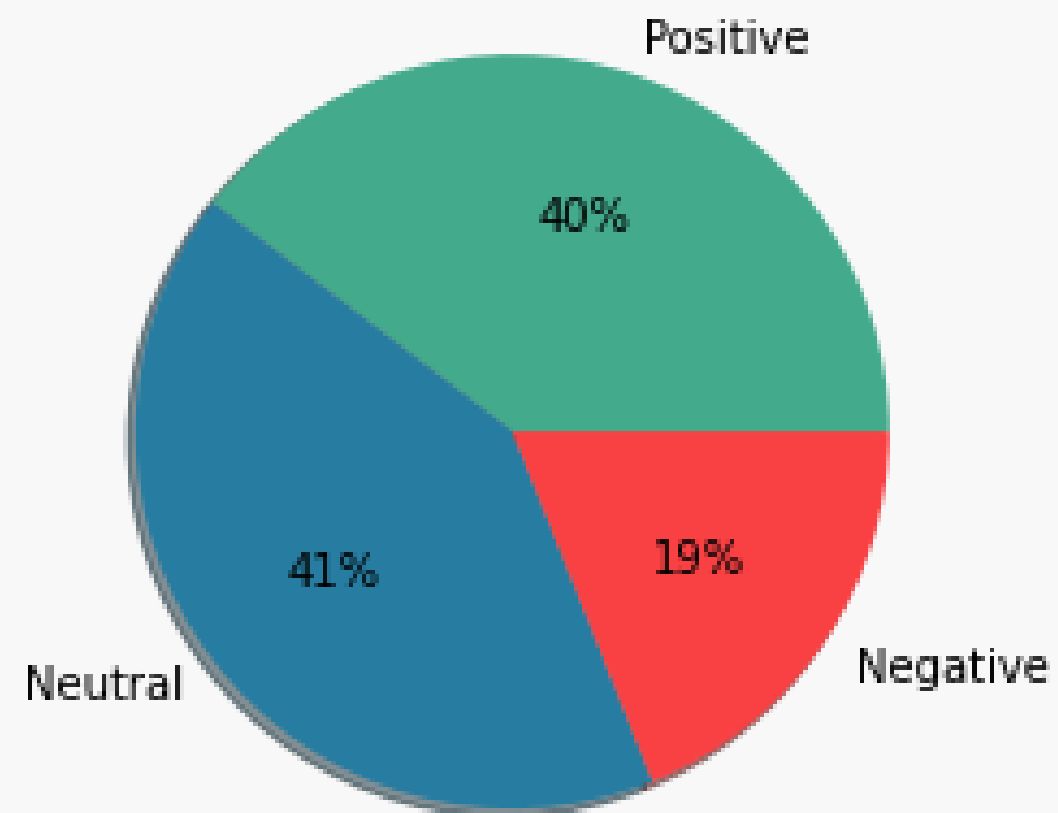
	tweet	sentiment
0	Cant believe were about to witness an Avengers Endgame level movie NEXT WEEK its been a while tbh	neutral
1	Im waiting	negative
2	One week away	neutral
3	Show me who you would cast as Harry Osborn in the MCU after	negative
4	New Scene from the streetfight for we see here a British Spiderman Sparking his marbles on the street something you never saw in original Spiderman Books cause that wouldnt work with anything spiderman does Nuts on Fire	positive
...	...	...
20750	Guess what month it is	neutral
20751	15 Days To Go	negative
20752	We know you could do better	positive
20753	Calling all you Spidey fans check out our exclusive combo available to pre order online during ticket purchase or in cinema from 15 December	positive
20754	Thank you for the update I wasn t sure if I was going to be able to participate in Nick s event since I had already secured coveted premier night tickets at my local happening that same night But now I can enjoy BOTH	positive
20755 rows × 2 columns		



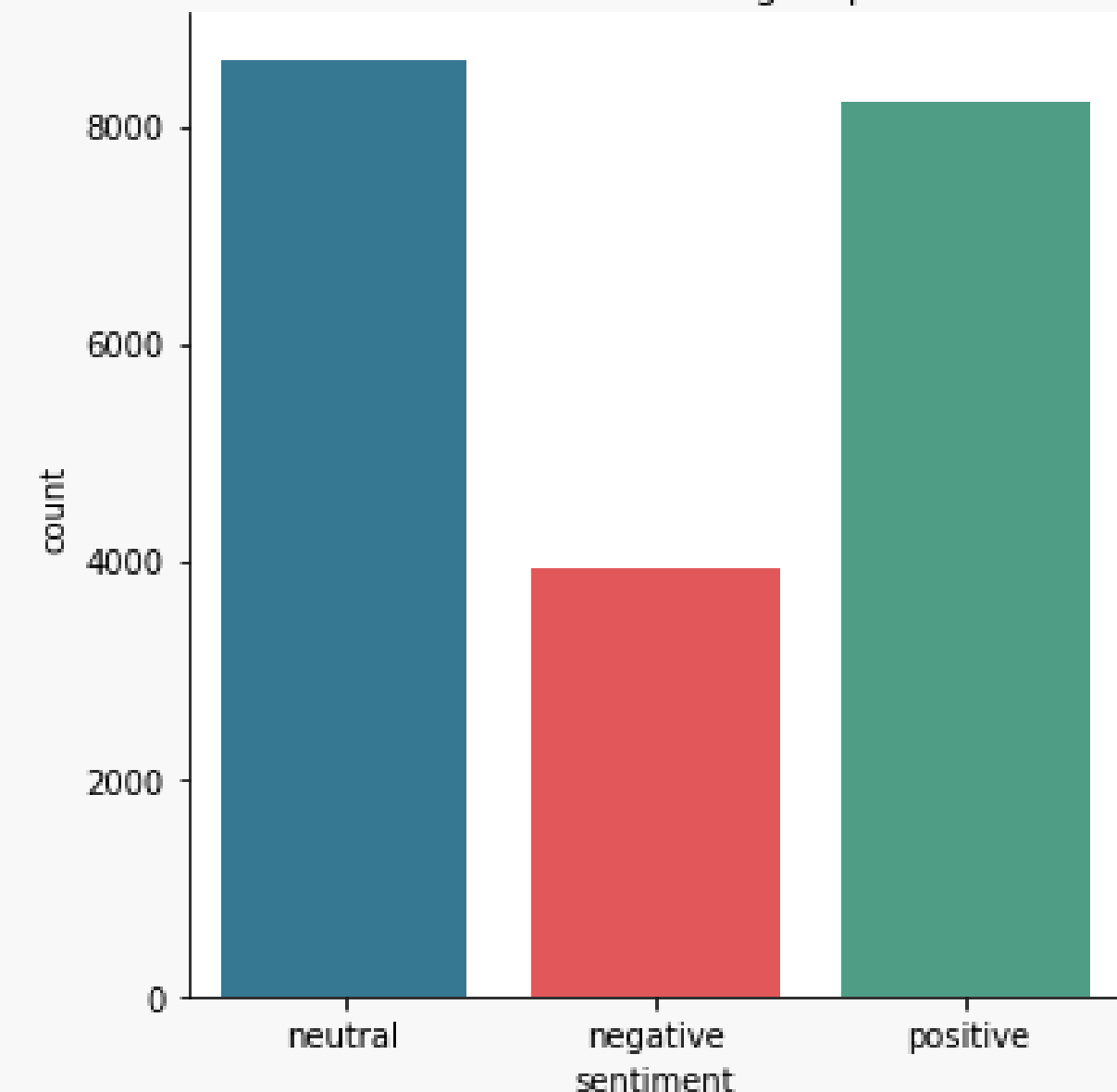
# Classification Result (Graph)

Berikut merupakan grafik **piechart** dan **barplot** dari **jumlah sentimen** pada **tweet** dengan hashtag **#SpiderManNoWayHome**.

Sentiment of Tweets with Hashtag #SpiderManNoWayHome



Sentiment of The Tweets with Hashtag #SpiderManNoWayHome



# Cross Validation

Untuk mengevaluasi kinerja dari model yang kita miliki, kita akan melakukan **Cross Validation**. Untuk fungsi **Cross Validation** ini, sudah saya masukkan pada fungsi **predict\_sentiments()**. Berikut kode untuk melakukan **Cross Validation**.



```
scores = cross_val_score(SVCmodel, textdata, sentiments, cv=5) # Do cross validation
print("Cross Validation Score : {}".format(scores))
```

## Hasil Cross Validation

```
Cross Validation Score : [0.9086967  0.90628764 0.913033   0.90773308 0.90146953]
```

Dari hasil tersebut, didapat **rata-rata akurasi** yang kita dapatkan sudah mencapai **90%**. Ini menunjukkan **prediksi sentimen** yang kita lakukan sudah cukup baik akurasi.

# Conclusion

Kesimpulan yang didapat dari **sentiment analysis** ini adalah dengan film **Spiderman No Way Home** ini memiliki sentimen yang didominasi **sentimen netral** dengan persentase **41%** dan disusul dengan **sentimen positif** dan **negatif** pada persentase masing-masing **40%** dan **19%** menunjukkan bahwa film **Spiderman No Way Home** ini memiliki **sentimen** yang **cenderung netral menuju positif**. Dengan hasil sentimen ini, patut ditunggu film tersebut untuk rilis dan disaksikan di bioskop.



# Reference

- <https://www.kaggle.com/stoicstatic/twitter-sentiment-analysis-for-beginners#TF-IDF-Vectoriser>
- <https://predictivehacks.com/how-to-get-a-sentiment-score-for-words-in-python/>
- <https://yunusmuhammad007.medium.com/tf-idf-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda>
- <https://dltsierra.medium.com/algoritma-tf-idf-633e17d10a80>

**Thank You**

**Thank You**

**Thank You**

**Thank You**