



PT. GIT SOLUTION

Prediksi Pergerakan Saham Menggunakan Sentiment Analysis

Kristal Belinda

Manajer Humas, Tim Humas



Forecasting Stock using NLP

Data Collection

- Saham yang digunakan adalah saham BBCA
- Data Saham diperoleh dari website Yahoo Finance
- Data Berita diperoleh dari Web Scrapping
- Data Twitter diperoleh dari proses Crawling

	datetime	headline_y	news_y	sentimen
0	2021-04-11	NaN	NaN	0.0
1	2021-04-12	IHSG anjlok 1,34% ke 5.989 pada sesi I hari in...	Reporter: Anna Suci Perwitasari Editor: Anna...	0.0
2	2021-04-13	IHSG dibuka melemah di awal perdagangan Selasa...	Reporter: Herlina KD Editor: Herlina Kartika...	0.0
3	2021-04-14	IHSG melonjak ke 6.013 pada akhir sesi I hari ...	Reporter: Anna Suci Perwitasari Editor: Anna...	1.0
4	2021-04-15	Saham big cap: IHSG bangkit, BBCA, BBRI, ARTO ...	Reporter: Khomarul Hidayat Editor: Khomarul ...	1.0
...
85	2021-07-05	PPKM Darurat! Ini Jam Buka BRI, BCA, BNI, Mand...	Jakarta, CNBC Indonesia - Pemberlakuan Pembata...	1.0
86	2021-07-06	Diskon PPnBM Diperpanjang, Cek Bunga-Syarat Kr...	Jakarta, CNBC Indonesia - Pemerintah kembali m...	1.0
87	2021-07-07	Buka-bukaan Bos OJK-Presdir BCA soal Nasib Kre...	Jakarta, CNBC Indonesia - Tingginya kasus Covi...	0.0
88	2021-07-08	IHSG menguat ke 6.050 pada perdagangan pagi in...	Reporter: Anna Suci Perwitasari Editor: Anna...	0.0
89	2021-07-09	Fakta-fakta di Balik Fenomena Ramai Leasing Gu...	Jakarta, CNBC Indonesia - Sebanyak 12 perusaha...	0.0
90 rows x 4 columns				

DATA BERITA

	Datetime	Tweet Id	Text	Username
0	2022-04-01 07:30:56+00:00	1509795431496433668	Tips\n•\nGempa\nMbanking BCA\nMuhammadiyah\n#S...	kalopsia50
1	2022-04-01 07:30:48+00:00	1509795400043343874	Digadang gantikan BCA Mobile,Aplikasi myBCA M...	suarapanturacom
2	2022-04-01 07:30:34+00:00	1509795338651340808	Inpo Mbanking Bca mu opo yo eror .?	andreasirwanto1
3	2022-04-01 07:30:25+00:00	1509795302982979586	BCA Mobile Ditutup? Apa Pengganti BCA Mobile? ...	suarapanturacom
4	2022-04-01 07:29:41+00:00	1509795117007511553	ternyata mbanking bca emg error pantesan td ma...	kesayangananon
...
7729	2022-01-02 04:39:09+00:00	1477499680686702594	Mbanking BCA Ulnya kalah modern sama BRI hahah...	multiakgae
7730	2022-01-02 03:47:45+00:00	1477486745126064131	@yteeam mbanking bca itu paling gampang and ga...	icapdf
7731	2022-01-02 02:08:44+00:00	1477461828749856768	Udh terbiasa sama pelayanan BCA dan pake mbank...	Nadagtx
7732	2022-01-01 22:14:09+00:00	1477402794210754563	Halobca cara ganti pin mbanking bca mobile saya	bella08579822
7733	2022-01-01 13:50:49+00:00	1477276124270919681	mbanking bca error apa gimana sih ih kesel bgt...	kthlof
7734 rows x 4 columns				

DATA TWITTER

Forecasting Stock using NLP

Labelling Data

Labelling data dilakukan menggunakan *Lexicon Model*

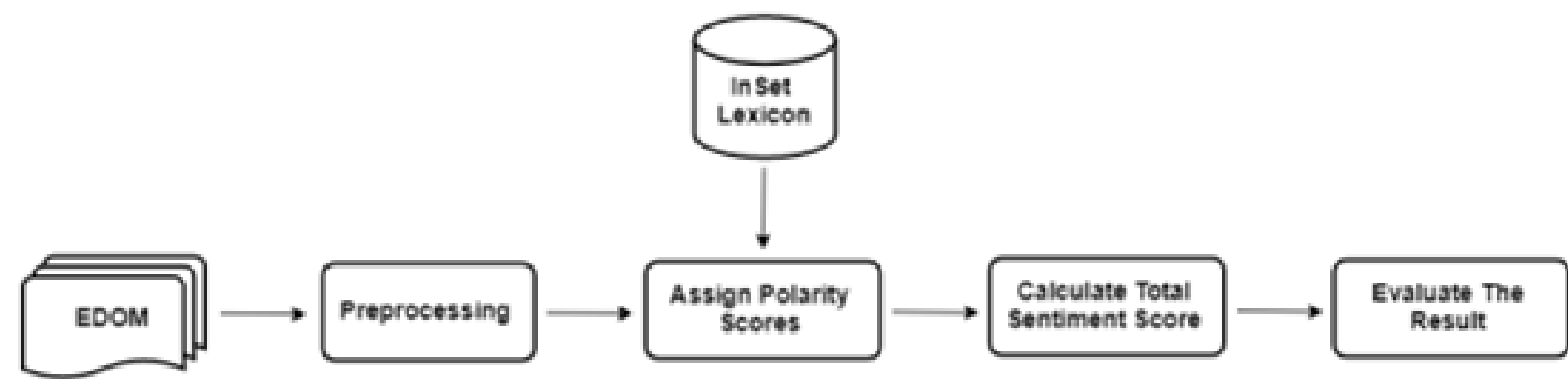


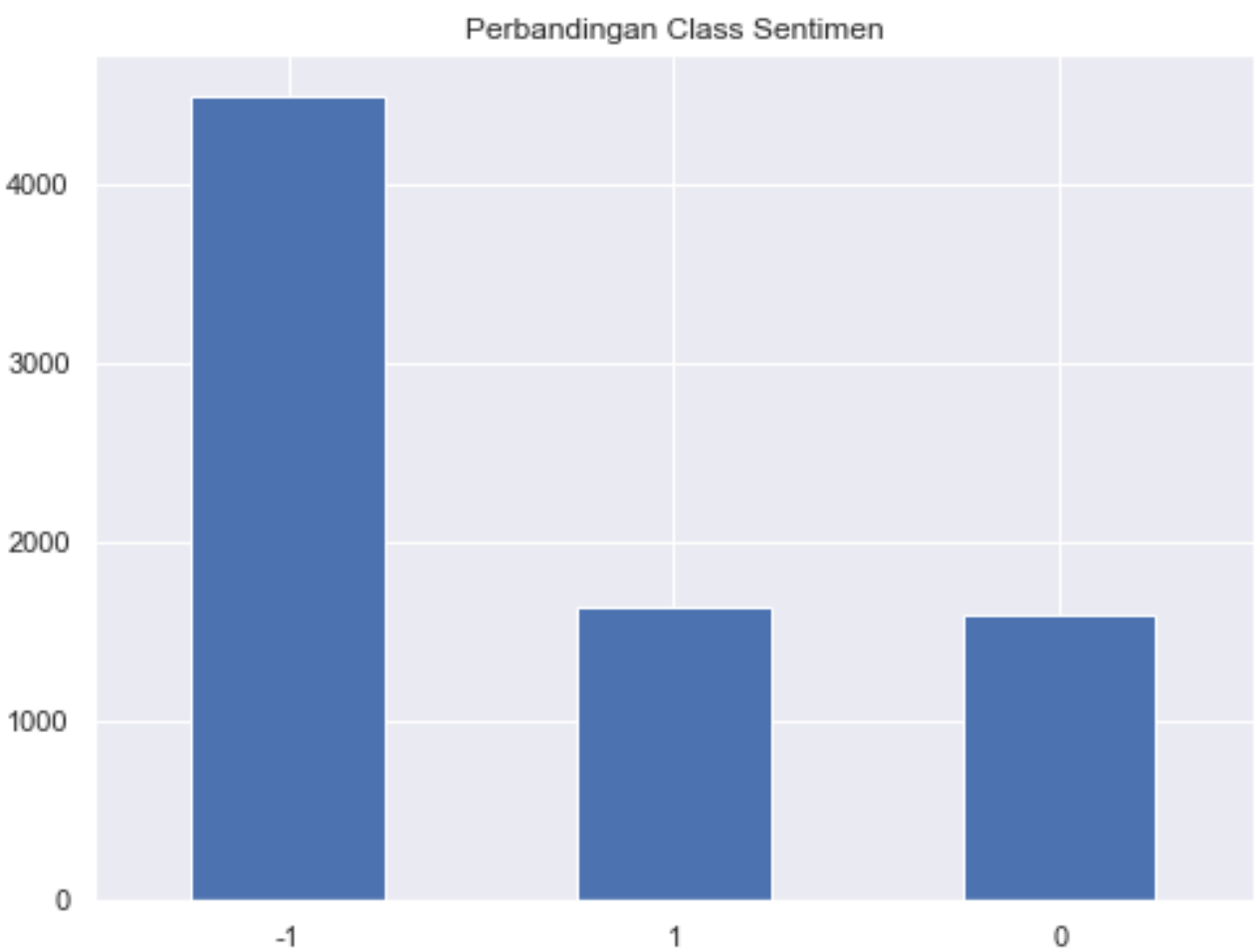
Fig. 1. Research Flowchart

Tahap Labelling Lexicon

TABLE V
RESULT OF SENTIMENT SCORE

Sentiment Score					
Sentiment words	=	'enak' (good)	'mudah'(easy)	'erti' (understand)	
Sentiment score	=	1	+	3	+ 0 = 4

Contoh Labelling



Perbandingan Class hasil Labelling

Forecasting Stock using NLP

Dataset

- Data saham yang digunakan 3 bulan (1 januari 2022 – 1 april 2021) dari yahoo finance
- Terdapat hari yang tidak terecord data nya
- Perlu labelling dari setiap data

	Date	Open	High	Low	Close	Adj Close	Volume
0	2022-01-01	7325.0	7400.0	7300.0	7325.0	7214.434082	54287400.0
1	2022-01-02	7325.0	7400.0	7300.0	7325.0	7214.434082	54287400.0
2	2022-01-03	7325.0	7400.0	7300.0	7325.0	7214.434082	54287400.0
3	2022-01-04	7325.0	7450.0	7325.0	7400.0	7288.301758	70624000.0
4	2022-01-05	7450.0	7525.0	7375.0	7450.0	7337.547363	76164900.0
...
86	2022-03-28	7825.0	7900.0	7800.0	7900.0	7900.000000	83911500.0
87	2022-03-29	7875.0	7900.0	7825.0	7850.0	7850.000000	62180600.0
88	2022-03-30	7900.0	7900.0	7825.0	7875.0	7875.000000	71430600.0
89	2022-03-31	7925.0	8000.0	7875.0	7975.0	7975.000000	112691900.0
90	2022-04-01	8075.0	8075.0	7850.0	7850.0	7850.000000	54491800.0
91 rows × 7 columns							

Dataset Saham

Forecasting Stock using NLP

Interpolasi Linear

Pada hari hari tertentu tidak tersedia record untuk harga saham, sehingga perlu diisi data kosong tersebut menggunakan metode interpolasi agar data harian terisi penuh.

BBCA Stock Price January - April 2022



Grafik Saham sebelum di interpolasi

BBCA Stock Price January - April 2022



Grafik saham setelah di Interpolasi

Forecasting Stock using NLP

Data Discretization

Semua data akan dikurangi dengan harga Close pada hari tersebut, jika harga lebih dari harga **Close** maka akan diganti dengan **Positive**, jika harga kurang dari harga Close akan diganti dengan **Negative**, jika sama dengan harga Close maka akan diganti dengan **Equal**.

	Date	Open	High	Low	Close	label
0	2022-01-01	Positive	Positive	Equal	7325.0	1
1	2022-01-02	Equal	Positive	Negative	7325.0	1
2	2022-01-03	Equal	Positive	Negative	7325.0	1
3	2022-01-04	Equal	Positive	Equal	7400.0	1
4	2022-01-05	Positive	Positive	Negative	7450.0	1
...
86	2022-03-28	Negative	Negative	Negative	7900.0	0
87	2022-03-29	Negative	Equal	Negative	7850.0	0
88	2022-03-30	Positive	Positive	Negative	7875.0	1
89	2022-03-31	Positive	Positive	Equal	7975.0	1
90	2022-04-01	Positive	Positive	Negative	7850.0	0
91 rows × 6 columns						

data setelah discretization

Combine Data

Menggabungkan Data Saham dan Teks berdasarkan Hari

	Date	Close	label	open_Equal	open_Positive	open_Negative	High_Equal	High_Positive	High_Negative	Low_Equal	Low_Negative	Low_Positive	sentimen
0	2022-04-02	7900.000000	0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0
1	2022-04-03	7900.000000	1	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1
2	2022-04-04	7900.000000	1	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0
3	2022-04-05	7900.000000	1	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0
4	2022-04-06	7750.000000	0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1
5	2022-04-07	7750.000000	1	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1
6	2022-04-08	7850.000000	1	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1
7	2022-04-09	7808.333333	0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0
8	2022-04-10	7766.666667	0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1
9	2022-04-11	7725.000000	0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0
10	2022-04-12	7800.000000	1	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0
11	2022-04-13	7800.000000	1	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0
12	2022-04-14	7700.000000	0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1

Data Setelah Digabungkan

Forecasting Stock using NLP

Evaluasi Model

Data di bagi menjadi data train dan data test dengan proporsi **8:2**

Accuracy	Recall	Precision	F1-Score
78.9 %	73 %	88 %	79 %

Kesimpulan

- Data twitter lebih menggambarkan sentimen dari pada kalimat dari berita
- Algoritma KNN cukup baik untuk memprediksi saham menggunakan data historikal dan sentimen, jika dilihat dari semua metrics yangimbang
- Diperlukan lebih dari satu tweet untuk menentukan sentimen dalam 1 hari

**Thank
You**

