

Pengenalan Pola Berbasis OCR untuk Pengambilan Data Bursa Saham

M. Dyovan Uidy Okta, Suci Aulia, dan Burhanuddin
Telkom University
Jl. Telekomunikasi Terusan Buah Batu, Bandung, 40257
e-mail: suciaulia@telkomuniversity.ac.id

Abstrak—Para investor dalam jual beli saham harus dapat menganalisa kapan secara tepat menjual dan membeli saham berdasarkan insting. Bagi para investor pemula, hal ini tentu saja menjadi kelemahan selain ketidaktepatan keputusan juga waktu yang dibutuhkan untuk menganalisa sejumlah data tidak efektif. Sehingga dibutuhkan suatu support-system yang dapat membantu para investor pemula untuk membuat keputusan dalam jual-beli saham. Support-system ini bekerja dengan membuat suatu tampilan kurva analisis secara online melalui data teks yang ada pada aplikasi harga saham Bursa Efek Indonesia (BEI) agar dapat dilakukan pengolahan data dengan prinsip pola pengenalan sehingga seterusnya dapat keluar suatu keputusan jual-beli yang dapat diperhitungkan untung ruginya oleh investor. Sebagai langkah awal dari sistem keseluruhan, pada penelitian ini telah dibangun suatu sistem konversi image to text berbasis Optical Character Recognition (OCR) yang dapat mengubah teks non-editable (.jpg) menjadi editable (.text) secara online. Setelah diperoleh data .text ini, kemudian akan digunakan pada penelitian selanjutnya untuk dilakukan analisis keputusan jual-beli saham. Berdasarkan hasil pengujian terhadap delapan perusahaan, diperoleh tingkat akurasi 96,8 % untuk konversi image to text berbasis OCR. Sedangkan untuk pengujian di Libre Office memiliki akurasi 100% dengan menggunakan font Droid serif, Takao PGhotic, dan Waree pada ukuran font 12pt.

Kata kunci: *stock data, character recognition, ocr, image to text, image processing*

Abstract—The investor must be able to use instinct to evaluate when to sell and buy stocks. This is, of fact, a weakness for inexperienced investors, in addition to the decision's inaccuracy and the time it takes to evaluate a slew of ineffective results. So that, a support system is needed to help the investors make decisions in buying and selling shares. This support system creates an online analysis curve display through text data in the BEI stock price application. The data processing based on pattern recognition will be carried out so that a buying and selling decision can be made to calculate the profit and loss by investors. As the first step of the whole system, this research has built an image-to-text conversion system based on OCR (Optical Character Recognition) that can convert the non-editable text (.jpg) to be editable (.text) online. After obtaining this .text data, the will used the system in further research to analyze stock buying and selling decisions. According to research on eight companies, the OCR-based image to text conversion has a 96.8% accuracy rate. Meanwhile, using Droid serif, Takao PGhotic, and Waree fonts at 12pt font sizes, it has 100 percent accuracy in Libre Office.

Keywords: *stock, character recognition, ocr, image to text, image processing*

I. PENDAHULUAN

Salah satu tempat jual beli saham di Indonesia adalah *Indonesia Stock Exchange (IDX)*. IDX bersifat *online trading* dengan perusahaan *broker*. Adapun aplikasi yang ada pada IDX untuk membantu para investor menganalisa perdagangan saham adalah Bursa Efek Indonesia (BEI). Namun, keluaran dari BEI bukanlah suatu keputusan untuk menjual atau membeli saham, tetapi ketersediaan data berupa dari fluktuasi saham dimana para investor harus menganalisa secara manual kapan secara tepat harus menjual dan membeli saham berdasarkan insting. Bagi para investor pemula, hal ini tentu saja menjadi kelemahan selain ketidaktepatan keputusan juga waktu

yang dibutuhkan untuk menganalisa sejumlah data tidak efektif. Sehingga dibutuhkan suatu *support-system* yang dapat membantu para investor pemula untuk membuat keputusan dalam jual-beli saham.

Penelitian ini merupakan tahap pertama dari keseluruhan sistem analisa jual beli saham otomatis berbasis *Optical Character Recognition (OCR)* [1]. OCR sebagai bagian dari *image processing* adalah teknik menerjemahkan karakter teks yang ditulis tangan, diketik, atau dicetak menjadi teks yang dikodekan mesin [2]. Penerapan OCR pada penelitian ini adalah untuk mengambil citra data harga saham pada aplikasi BEI kemudian mengkonversinya menjadi data berupa *text-editable*. Dengan perolehan data teks *editable* dari BEI tersebut, pada penelitian tahap kedua

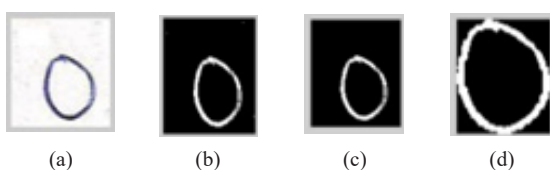
dapat dilakukan pengolahan data dengan membuat grafik fluktuasi kemudian melakukan pola pengenalan untuk membantu dalam menganalisa kemungkinan kerugian atau keuntungan dari perubahan harga saham tersebut. Pada penelitian ini, permasalahan yang akan dibahas terdiri dari proses akuisisi *screen capturing* dengan interval tertentu dari PT. Indo Premier, proses *text recognition*, proses *text-parsing*, kemudian proses penyimpanan hasil *parsing* ke *database*. Penelitian tahap-1 ini kemudian akan digabungkan dengan penelitian tahap-2, yaitu sistem analisis jual beli saham berdasarkan data pada tahap-1 berbasis *Moving Average* (MA). Pada tahap-2, data akan dibuat berupa kurva, kemudian akan dilatih pengenalan polanya berbasis *machine learning*. Sehingga seterusnya dapat keluar suatu keputusan jual-beli yang dapat diperhitungkan untung ruginya oleh investor dalam memprediksikan fluktuasi harga saham.

OCR banyak digunakan sebagai deteksi plat nomor kendaraan [3-5], deteksi uang palsu [6], formulir pemasukan data berupa tulisan tangan dalam berbagai bahasa [7-11]. OCR dapat mengenali text dalam berbagai bahasa dengan tingkat *error* rata-rata 15 % seperti dalam penjelasan *survey* yang dilakukan oleh Singh [12] bahkan sampai konversi *image to text* dari huruf arab gundul [13] dan kanji [14], dan rata-rata tingkat akurasi di atas 80% hasil *survey* perbandingan berbagai metodologi dan *classifier* [15-18] OCR secara *online* telah berhasil disimulasikan oleh Onyejebu berbasis *cloud azure* [19]. Pada penelitian sebelumnya, analisa jual beli saham dilakukan berdasarkan pola sinyal yang dihasilkan oleh *Indikator Commodity Channel Index* (CCI) [20] berdasarkan pendekatan *Rule-Based Evidential Reasoning* dan metode *Fuzzy C-Means* (FCM), namun hasilnya masih membutuhkan verifikasi. Maka pada penelitian ini diusulkan suatu solusi dengan pemanfaatan teknologi OCR, suatu *support-system* yang dapat membantu para investor pemula untuk membuat keputusan dalam jual-beli saham.

II. METODE

A. Pengolahan Citra dan OCR

Tujuan dari sistem OCR adalah untuk mengubah gambar dokumen yang dipindai dalam format cetak atau *print* ke dalam format digital agar datanya dapat dicari dan diedit [21]. Proses dari transformasi OCR ditunjukkan pada Gambar 1, dan ilustrasi hasil konversinya pada Gambar 2. Pada Gambar 1, diilustrasikan proses binerisasi sebagai bagian dari *preprocessing* OCR dengan tujuan untuk memudahkan proses segmentasi karakter [22][23].



Gambar 1. Ilustrasi pengolahan citra (a) segmentasi karakter, (b) binerisasi, (c) pengurangan noise, (d) penebalan karakter.



Gambar 2. Ilustrasi image to text hasil OCR

B. Segmentasi Citra dan Pytesseract

Proses mempartisi citra menjadi beberapa daerah atau objek disebut dengan Segmentasi. Segmentasi citra biasanya berdasar pada sifat *discontinuity* atau *similarity* dari intensitas piksel [24][26]. Secara umum proses segmentasi karakter dengan OCR dapat dilihat pada Gambar 3.

Pytesseract atau *python tesseract* adalah alat OCR untuk python, *Pytesseract* berfungsi untuk mengenali dan membaca text tertanam dalam suatu gambar. Python tesseract merupakan pembungkus untuk Google Tesseract-OCR dimana sudah mendukung untuk melakukan pembacaan jenis gambar yang telah didukung oleh *Python Imaging Library* (PIL) termasuk jpeg,png,bmp,tiff, dan lain lain. Dimana Tesseract-OCR secara *default* hanya mendukung gambar dengan jenis format bmp dan tiff.

C. OCR-Tesseract

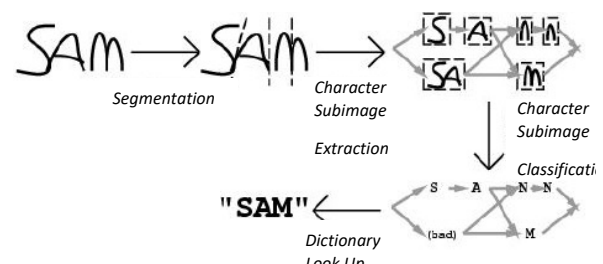
OCR-Tesseract merupakan paket yang berisi mesin OCR `-lib tesseract` dan baris program `-tesseract`. Pengembang utama Tesseract adalah Ray Smith yang kemudian dikembangkan oleh Zdenko Podobny[25]. Tesseract memiliki dukungan *unicode* UTF-8 dan dapat mengenali lebih dari 100 bahasa diluar kotak. Hal ini dilakukan untuk mengenali bahasa lain dengan keluaran berformat *plain text*, *hocr* (html), pdf. Contoh hasil keluaran Tesseract diilustrasikan pada Gambar 4.

Untuk mengukur performansi sistem, setelah dilakukan pengujian akan dihitung dengan menggunakan persamaan 1 berikut.

$$Akurasi = \frac{\sum \text{data benar}}{\sum \text{total data}} \quad (1)$$

III. ARSITEKTUR DAN PERANCANGAN SISTEM

Fungsi umum sistem terbagi menjadi dua bagian. Pertama adalah sebagai pengolah gambar yang diterima melalui pengambilan gambar di *website broker* lalu

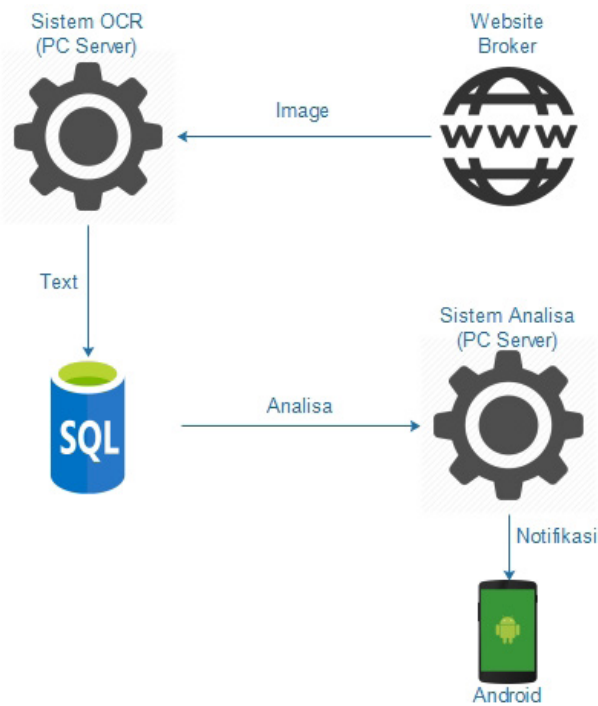


Gambar 3. Segmentasi karakter pada citra

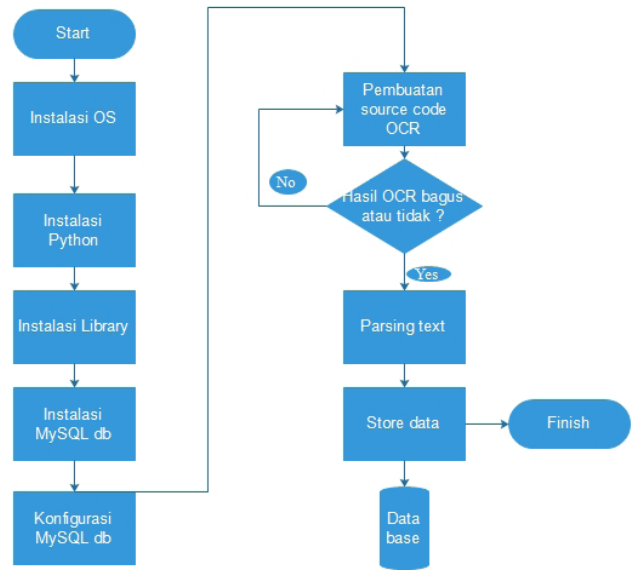
Original	Thresholded	OCR
		66htv
		5n7pf
		qvsxp
		6x94d
		jmsq
		tcédb

Gambar 4. Ilustrasi penggunaan OCR-tesseract

mengkonversinya menjadi *string*, dan kemudian dilakukan *text parsing* untuk dipilah data-data yang diinginkan kemudian dilakukan penyetoran ke *SQL Hosting*. Sedangkan sistem kedua melakukan analisis data-data yang sebelumnya telah dikumpulkan dari sistem pertama di dalam *database* dan melakukan pengiriman sinyal atau notifikasi ke android. Secara sistematis, infrastruktur sistem keseluruhan yang dirancang dan akan diimplementasikan memiliki topologi seperti pada Gambar 5.



Gambar 5. Gambaran umum aplikasi sistem analisa fluktuasi saham berbasis OCR pada android.



Gambar 6. Diagram alir perancangan sistem

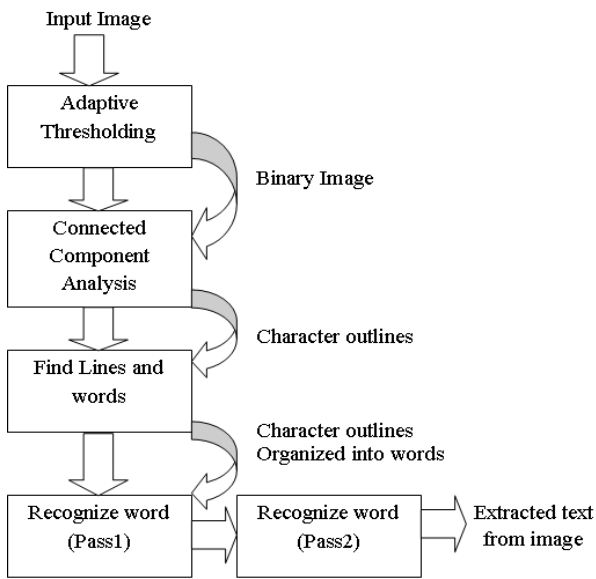
Penelitian ini merupakan tahap pertama dari sistem keseluruhan seperti pada Gambar 5. Adapun diagram alir penelitian ini dijabarkan pada Gambar 6.

Berdasarkan Gambar 6, sistem OCR berfungsi sebagai media untuk menerjemahkan gambar yang diperoleh dari *website broker* menjadi data berupa *String* atau *Text* yang akan dikirim ke *SQL hosting*. *Website broker* memiliki fungsi sebagai penyedia informasi berupa fluktuasi harga ataupun transaksi jual beli saham secara *realtime*. *SQL hosting* merupakan tempat penyimpanan *database* secara *online* agar dapat diakses oleh perangkat lain yaitu *mobile application*. Sistem Analisa berfungsi sebagai media untuk melakukan analisa data dari *database* sehingga menghasilkan keluaran berupa notifikasi. Aplikasi pada android berfungsi sebagai media penerima notifikasi dari sistem analisa sehingga *user* dapat mengetahui informasi dari notifikasi tersebut.

Pada penelitian ini, metode untuk pengambilan gambar melalui *website* yaitu dengan menggunakan *pyscreenshot* modul, dimana modul ini dapat mengambil *screenshot* keseluruhan layar *desktop* secara penuh ataupun pada suatu area tertentu dengan ukuran pixel dan koordinat yang diinginkan. Setelah proses *screenshot*, dilanjutkan dengan proses *cropping* menggunakan *Python Image Library* (PIL). Kemudian dilakukan segmentasi karakter pada hasil *cropping* tersebut dengan memanfaatkan Tesseract library di python untuk tujuan *text parsing*. Hasil *text parsing* yang masih dalam format *image* kemudian dikonversi ke dalam *string* dalam format *text* yang disimpan dalam *database* dengan menggunakan *Database MySQL*.

Tesseract OCR bekerja melalui langkah-langkah sesuai diagram alir yang ditunjukkan pada Gambar 7.

Langkah pertama adalah *Adaptive Treshold*, yaitu mengubah gambar menjadi gambar biner. Langkah selanjutnya adalah menghubungkan analisis komponen, yang digunakan untuk melakukan ekstraksi ciri setiap karakter. Sehingga hasilnya memisahkan karakter sebagai objek dengan bit "1" atau putih dan *background* dengan



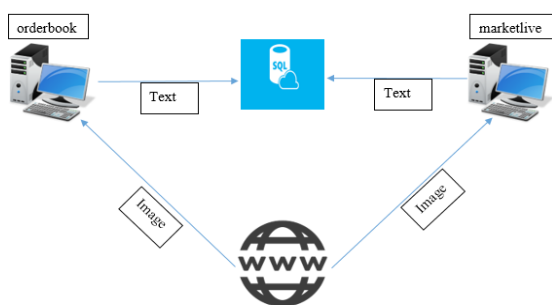
Gambar 7. Arsitektur Tesseract pada OCR

bit "0" atau hitam.

Kemudian setelah itu *character outlines* dikonversi menjadi *Binary Long Object* (BLOB). BLOB tersebut akan disusun menjadi baris teks, baris, dan daerah yang kemudian dianalisis untuk area dan teks yang setara. Kemudian teks dibagi menjadi kata dengan dipisahkan oleh spasi. Pada *Recognize Word* (Pass 1) terjadi proses *text recognition* yang akan diteruskan ke tahap *Adaptive Classifier*. *Adaptive Classifier* akan mengklasifikasikan teks yang telah dikenali sesuai *database* kemudian akan menampilkan karakter tersebut sebagai hasil konversi *image to text* oleh sistem.

IV. HASIL DAN PEMBAHASAN

Pada pengujian dan analisis, akan dilakukan pengujian dan analisa kemampuan dari dua buah sistem OCR. Sistem pertama akan melakukan *running* dan *image processing* pada menu *orderbook* dengan mengambil data-data berupa kode perusahaan, harga terakhir, harga sebelumnya, perubahan harga, perubahan harga dalam persen, harga buku dihari itu, harga tertinggi, harga terendah, jumlah lot, ekivalensi saham, dan harga rata-rata. Sistem kedua



Gambar 8. Ilustrasi pengujian sistem OCR

akan melakukan *running* dan *image processing* pada menu *livemarket* dengan mengambil data-data berupa waktu, kode perusahaan, harga, perubahan harga, perubahan harga dalam persen, jumlah lot, kode pembeli, kode penjual. Skenario pengujian kedua buah sistem OCR tersebut diilustrasikan pada Gambar 8.

A. Image Processing

Pada pengujian pertama akan dilakukan pengujian dengan melihat tingkat akurasi dari hasil OCR yaitu dengan cara melakukan pencocokan antara citra hasil *screenshot* dengan data yang telah dikirim ke *database*. Hasil pengambilan data *screenshot* dari *website* secara *online* secara layar penuh pada desktop PC ditunjukkan pada Gambar 9a. Gambar 9a merupakan hasil pengambilan *screenshot* secara langsung dari *website broker* Indopremier sekuritas. Disini ditampilkan delapan jenis perusahaan yang berbeda dengan menampilkan data-data berupa harga penawaran dan permintaan saham beserta jumlah lot nya. Data di atas akan selalu mengalami perubahan secara *realtime* selama waktu masih menunjukkan jam buka saham. Kemudian dilakukan *cropping* dengan perbesaran 133% pada Gambar 9a, hasilnya ditunjukkan pada Gambar 9b. Gambar 9b merupakan hasil *cropping* objek yang menunjukkan data saham pada delapan perusahaan, kemudian dilanjutkan dengan segmentasi dan *cropping* pada sebuah perusahaan dengan menghilangkan bagian-bagian yang tidak diperlukan dalam analisis data sebagai contoh hasilnya pada Gambar 9c.

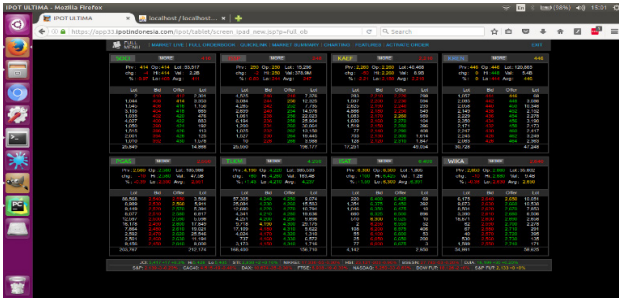
Proses *cropping* masih berlanjut pada setiap kelompok teks pada Gambar 9c. Gambar 9d (80 x 36 piksel) merupakan gambar hasil *cropping* dengan koordinat 114,101,152,120 dari Gambar 9c. Gambar 9e (38 x 19 piksel) merupakan gambar setelah dilakukan *resize* pada Gambar 9d.

B. Proses OCR

Pada proses *Recognize Word* (Pass 1) berusaha mengenali setiap kata dari teks. Setiap kata yang lulus dengan akurasi tinggi pada proses pelatihan akan diteruskan ke *Adaptive Classifier*. *Adaptive Classifier* telah menerima beberapa data pelatihan dan mempelajari sesuatu yang baru, sehingga dapat digunakan untuk menyelesaikan dalam mengambil teks dari gambar. Gambar 10 adalah teks hasil OCR yang ditampilkan dalam format .txt.

Dengan prosedur yang sama diujikan proses OCR pada tujuh perusahaan lainnya dengan hasil yang dapat dilihat pada Gambar 11 – Gambar 17.

Dari pengujian yang telah dilakukan didapatkan bahwa dari 320 baris data yang diujikan dengan OCR terdapat 10 baris data dengan hasil yang tidak sesuai dengan image aslinya. Yaitu kode perusahaan SOCI yang mana terbaca dalam OCR menjadi SOCI. Lalu dalam satu kali proses OCR delapan perusahaan di atas membutuhkan waktu 2 detik dalam melakukannya. Sehingga dapat disimpulkan bahwa presentasi akurasi dalam pengujian OCR di atas



(a)

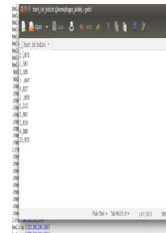
Prv: 414	Op: 414	Lot: 55,529
chg: -4	Hi: 414	Val: 2.3B
%: -0.97	Lo: 408	Avg: 411

Lot	Bid	Offer	Lot
1,078	408	410	1,439
1,583	406	412	2,301
3,105	404	414	3,353
1,047	402	416	1,150
4,027	400	418	665
1,050	398	420	476
1,515	396	422	883
2,001	394	424	192
1,010	392	426	113
4,000	390	428	125
25,931			16,305

(a)

SOCI	MORE	410	ISSP	MORE	248	KAEF	MORE	2,210	KREN	MORE	446
Prv: 414	Op: 414	Lot: 55,529	Prv: 250	Op: 250	Lot: 15,298	Prv: 2,280	Op: 2,280	Lot: 40,468	Prv: 446	Op: 446	Lot: 120,865
chg: -4	Hi: 414	Val: 2.3B	chg: -0	Hi: 250	Val: 379.0M	chg: -20	Hi: 2,280	Val: 8.9B	chg: 0	Hi: 444	Val: 5.4B
%: -0.97	Lo: 408	Avg: 411	%: -0.00	Lo: 244	Avg: 247	%: -2.21	Lo: 2,180	Avg: 2,210	%: 0	Lo: 444	Avg: 446

(b)



(b)

SOCI	1078	408	410	1439	2016-10-27	15:16:37
SOCI	1583	406	412	2301	2016-10-27	15:16:37
SOCI	3105	404	414	3353	2016-10-27	15:16:37
SOCI	1047	402	416	1150	2016-10-27	15:16:37
SOCI	4027	400	418	665	2016-10-27	15:16:37
SOCI	1050	398	420	476	2016-10-27	15:16:37
SOCI	1515	396	422	883	2016-10-27	15:16:37
SOCI	2001	394	424	192	2016-10-27	15:16:37
SOCI	1010	392	426	113	2016-10-27	15:16:37
SOCI	4000	390	428	125	2016-10-27	15:16:37

(c)

Gambar 10. (a) Gambar hasil cropping dari website, (b) Hasil OCR, (c) Hasil OCR setelah dilakukan parsing teks.

Prv: 414	Op: 414	Lot: 55,529
chg: -4	Hi: 414	Val: 2.3B
%: -0.97	Lo: 408	Avg: 411

Lot	Bid	Offer	Lot
1,078	408	410	1,439
1,583	406	412	2,301
3,105	404	414	3,353
1,047	402	416	1,150
4,027	400	418	665
1,050	398	420	476
1,515	396	422	883
2,001	394	424	192
1,010	392	426	113
4,000	390	428	125
25,931			16,305

(c)

408

(d)

408

(e)

Gambar 9. (a) Tampilan website broker, (b) Proses cropping, (c) Proses Segmentasi, (d) Hasil segmentasi dari c, (e) Hasil resize ratio.

yaitu 96.8 %.

Selain itu, sistem OCR diujikan dengan font yang berbeda seperti yang ditunjukkan pada Tabel 1. Salah satu hasil OCR sistem dengan variasi Font Waree ditunjukkan pada Gambar 18.

Selain itu variasi ukuran font diuji cobakan dengan mengganti font size pada 9 pt dan 14 pt. Hasil yang diperoleh memiliki tingkat kesalahan OCR yang lebih besar, sehingga dapat disimpulkan size optimal dalam OCR yaitu pada font 12 pt.

Berdasarkan hasil pengujian *image to text*, sistem

Prv: 250	Op: 250	Lot: 15,298
chg: -2	Hi: 250	Val: 379.0M
%: -0.80	Lo: 244	Avg: 247

Lot	Bid	Offer	Lot
4,525	246	248	7,374
3,084	244	250	12,325
4,265	242	252	7,735
2,659	240	254	14,976
1,061	238	256	22,023
6,184	236	258	25,904
1,200	234	260	30,064
1,025	232	262	13,150
1,527	230	264	19,445
10	226	266	3,988
25,550			196,175

(a)

ISSP	4525	246	248	7374	2016-10-27	15:16:37
ISSP	3084	244	250	12325	2016-10-27	15:16:37
ISSP	4265	242	252	7735	2016-10-27	15:16:37
ISSP	2659	240	254	14976	2016-10-27	15:16:37
ISSP	1061	238	256	22023	2016-10-27	15:16:37
ISSP	6194	236	258	25904	2016-10-27	15:16:37
ISSP	1200	234	260	30064	2016-10-27	15:16:37
ISSP	1025	232	262	13150	2016-10-27	15:16:37
ISSP	1527	230	264	19445	2016-10-27	15:16:37
ISSP	10	226	266	3988	2016-10-27	15:16:37

(b)

Gambar 11. (a) Gambar hasil cropping perusahaan ke-2 ISSP, (b) Hasil dari OCR

Prv: 2,280	Op: 2,280	Lot: 40,468
chg: -50	Hi: 2,280	Val: 8.9B
%: -2.21	Lo: 2,180	Avg: 2,210

Lot	Bid	Offer	Lot
283	2,210	2,220	299
1,597	2,200	2,230	602
2,625	2,190	2,240	233
4,866	2,180	2,250	549
1,063	2,170	2,260	989
1,600	2,160	2,270	104
1,519	2,150	2,280	396
77	2,140	2,290	408
703	2,130	2,300	1,614
128	2,120	2,310	1,847
17,251			49,062

(a)

KAEF	283	2210	2220	299	2016-10-27	15:16:38
KAEF	1597	2200	2230	602	2016-10-27	15:16:38
KAEF	2625	2190	2240	233	2016-10-27	15:16:38
KAEF	4866	2180	2250	549	2016-10-27	15:16:38
KAEF	1083	2170	2260	989	2016-10-27	15:16:38
KAEF	1600	2160	2270	104	2016-10-27	15:16:38
KAEF	1519	2150	2280	396	2016-10-27	15:16:38
KAEF	77	2140	2290	408	2016-10-27	15:16:38
KAEF	703	2130	2300	1614	2016-10-27	15:16:38
KAEF	128	2120	2310	1847	2016-10-27	15:16:38

(b)

Gambar 12. (a) Gambar hasil cropping perusahaan ke-3 KAEF, (b) Hasil dari OCR.

Prv: 446	Op: 446	Lot: 120,865
chg: 0	Hi: 444	Val: 5.4B
%: 0	Lo: 444	Avg: 446

Lot	Bid	Offer	Lot
1,057	444	446	69
2,085	442	448	3,086
2,058	440	450	10,346
2,149	438	452	2,152
2,229	436	454	2,278
2,350	434	456	3,190
2,171	432	458	2,173
2,247	430	460	2,417
2,245	428	462	3,249
2,063	426	464	2,363
30,728			47,248

(a)

KREN	1057	444	446	69	2016-10-27	15:16:38
KREN	2085	442	448	3086	2016-10-27	15:16:38
KREN	2058	440	450	10346	2016-10-27	15:16:38
KREN	2149	438	452	2152	2016-10-27	15:16:38
KREN	2229	436	454	2278	2016-10-27	15:16:38
KREN	2350	434	456	3190	2016-10-27	15:16:38
KREN	2171	432	458	2173	2016-10-27	15:16:38
KREN	2247	430	460	2417	2016-10-27	15:16:38
KREN	2245	428	462	3249	2016-10-27	15:16:38
KREN	2063	426	464	2363	2016-10-27	15:16:38

(b)

Gambar 13. (a) Gambar hasil cropping perusahaan ke-4 KREN, (b) Hasil dari OCR.

Lot	Bid	Offer	Lot
88,522	2,580	2,550	4,579
8,988	2,530	2,560	5,911
9,449	2,520	2,570	4,450
8,877	2,510	2,580	6,317
12,597	2,500	2,590	5,588
16,178	2,490	2,600	17,849
7,854	2,480	2,610	18,921
2,592	2,470	2,620	25,546
2,801	2,460	2,630	11,194
9,458	2,450	2,640	9,900
204,521			212,241

PGAS	88522	2540	2550	4579	2016-10-27	15:16:37
PGAS	8989	2530	2560	5911	2016-10-27	15:16:37
PGAS	9449	2520	2570	4450	2016-10-27	15:16:37
PGAS	8877	2510	2580	6317	2016-10-27	15:16:37
PGAS	12597	2500	2590	5588	2016-10-27	15:16:37
PGAS	16178	2490	2600	17849	2016-10-27	15:16:37
PGAS	7864	2480	2610	19021	2016-10-27	15:16:37
PGAS	2592	2470	2620	25546	2016-10-27	15:16:37
PGAS	2501	2460	2630	11194	2016-10-27	15:16:37
PGAS	9456	2450	2640	9000	2016-10-27	15:16:37

(a) (b)

Gambar 14. (a) Gambar hasil cropping perusahaan ke-5 PGAS, (b) Hasil dari OCR.

Lot	Bid	Offer	Lot
57,261	4,240	4,250	10,187
25,085	4,230	4,260	15,578
12,890	4,220	4,270	10,794
4,341	4,210	4,280	19,636
4,251	4,200	4,290	9,898
9,718	4,190	4,300	28,671
17,109	4,180	4,310	5,622
4,024	4,170	4,320	1,310
737	4,160	4,330	6,572
3,173	4,150	4,340	1,716
168,357			157,314

TLKM	57261	4240	4250	10187	2016-10-27	15:16:38
TLKM	25085	4230	4260	15578	2016-10-27	15:16:38
TLKM	12690	4220	4270	10794	2016-10-27	15:16:38
TLKM	4341	4210	4280	19794	2016-10-27	15:16:38
TLKM	4251	4200	4290	9898	2016-10-27	15:16:38
TLKM	9718	4190	4300	28671	2016-10-27	15:16:38
TLKM	17109	4180	4310	5622	2016-10-27	15:16:38
TLKM	4024	4170	4320	1310	2016-10-27	15:16:38
TLKM	737	4160	4330	6572	2016-10-27	15:16:38
TLKM	3173	4150	4340	1716	2016-10-27	15:16:38

(a) (b)

Gambar 15. (a) Gambar hasil cropping perusahaan ke-6 TLKM, (b) Hasil dari OCR.

Lot	Bid	Offer	Lot
220	6,400	6,425	60
1,354	6,375	6,450	392
1,046	6,350	6,475	10
660	6,325	6,500	895
510	6,300	6,525	10
2	6,250	6,550	52
108	6,200	6,575	406
55	6,100	6,600	53
25	6,050	6,650	200
77	6,000	6,675	3
4,142			2,950

ISAT	220	6400	6425	60	2016-10-27	15:16:37
ISAT	1354	6375	6450	392	2016-10-27	15:16:37
ISAT	1046	6350	6475	10	2016-10-27	15:16:37
ISAT	660	6325	6500	896	2016-10-27	15:16:37
ISAT	510	6300	6525	10	2016-10-27	15:16:37
ISAT	2	6250	6550	52	2016-10-27	15:16:37
ISAT	108	6200	6575	406	2016-10-27	15:16:37
ISAT	55	6100	6600	53	2016-10-27	15:16:37
ISAT	25	6050	6650	200	2016-10-27	15:16:37
ISAT	77	6000	6675	3	2016-10-27	15:16:37

(a) (b)

Gambar 16. (a) Gambar hasil cropping perusahaan ke-7 ISAT, (b) Hasil dari OCR.

Lot	Bid	Offer	Lot
6,127	2,640	2,650	10,366
9,879	2,630	2,660	10,538
5,004	2,620	2,670	7,853
3,390	2,610	2,680	6,506
18,871	2,600	2,690	2,858
62	2,590	2,700	2,278
87	2,580	2,710	231
40	2,570	2,720	395
530	2,560	2,730	135
1,599	2,550	2,740	171
51,448			58,940

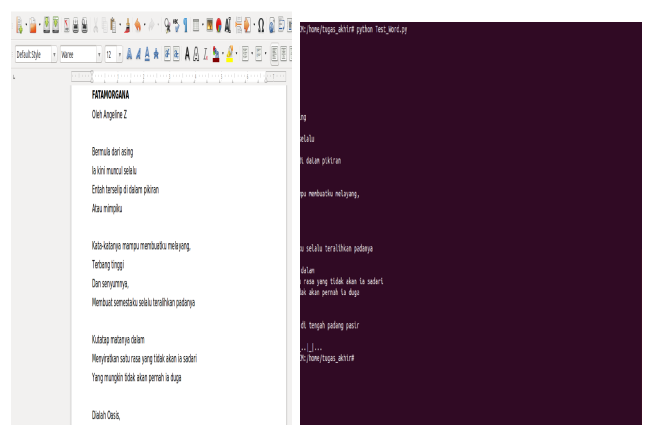
WIKA	6127	2640	2650	10366	2016-10-27	15:16:38
WIKA	9879	2630	2660	10538	2016-10-27	15:16:38
WIKA	5004	2620	2670	7853	2016-10-27	15:16:38
WIKA	3390	2610	2680	6506	2016-10-27	15:16:38
WIKA	18871	2600	2690	2658	2016-10-27	15:16:38
WIKA	62	2590	2700	2278	2016-10-27	15:16:38
WIKA	67	2580	2710	291	2016-10-27	15:16:38
WIKA	40	2570	2720	395	2016-10-27	15:16:38
WIKA	530	2560	2730	135	2016-10-27	15:16:38
WIKA	1599	2550	2740	171	2016-10-27	15:16:38

(a) (b)

Gambar 17. (a) Gambar hasil cropping perusahaan ke-8 ISAT, (b) Hasil dari OCR.

Tabel 1. Pengujian OCR dengan font yang berbeda

No	Font	Tingkat akurasi		
		9pt	12pt	14pt
1.	Comic Sains MS	47%	73%	64%
2.	Droid Serif	61%	100%	83%
3.	Liberation Serif	82%	94%	93%
4.	Takao P Gothic	78%	100%	95%
5.	Times New Roman	72%	83%	83%
6.	Verdana	79%	83%	79%
7.	Waree	96%	100%	100%
8.	Arial	86%	96%	96%



(a) (b)

Gambar 18. (a) Pengujian libre office dengan font Waree, (b) Hasil OCR

berhasil mengkonversi dengan tingkat akurasi 100% pada tiga font (12pt) saja, yaitu font Droid Serif, Takao P Gothic, dan Waree. Sehingga, pada penelitian selanjutnya ketiga font ini yang akan digunakan pada tahap kedua dari keseluruhan sistem. Keseluruhan sistem yang akan dibangun adalah sistem yang mampu menganalisis perdagangan jual-beli saham secara otomatis. Sistem ini bekerja berdasarkan masukkan data saham secara *real-time* dari beberapa perusahaan yang menjadi objek studi pada penelitian tahap pertama ini. Penelitian tahap kedua bertugas menghitung masing-masing tiga nilai *Mean Average* (MA) perusahaan. Nilai-nilai tersebut akan dijadikan acuan analisis yang berbasis *machine learning* menggunakan metoda *Moving Average Convergence Divergence* (MACD) dan *support resistance* [27].

V. KESIMPULAN

Dalam studi ini telah dirancang dan diimplementasikan sistem konversi *image-to-text* berbasis OCR pada aplikasi BEI melalui *screen capturing* dengan interval tertentu dari PT. Indo Premier. Berdasarkan hasil pengujian terhadap delapan perusahaan, diperoleh tingkat akurasi 96,8 % untuk konversi *image-to-text* berbasis OCR dengan waktu proses konversi rata-rata sebesar 2 detik. Sedangkan untuk pengujian jenis teks di libre office memiliki akurasi 100% dengan menggunakan font Droid serif, Takao P Gothic, dan Waree pada ukuran font 12pt. Dengan adanya data hasil konversi *image-to-text* pada aplikasi BEI ini, diharapkan dapat digunakan pada penelitian selanjutnya untuk dibuat dalam bentuk grafik fluktuasi. Dari grafik fluktuasi tersebut kemudian dilakukan pengenalan pola atau *pattern recognition* untuk membantu dalam menganalisa kemungkinan kerugian atau keuntungan dari perubahan harga saham tersebut.

REFERENSI

[1] A. Chaudhuri, K. Mandaviya, P. Badelia, and S. K Ghosh, "Optical character recognition systems for different languages with soft computing," *Springer*, vol. 352, no. December, pp. 260,

- 2017.
- [2] M. K. Sahu and N. K. Dewangan, "Handwritten character recognition using neural network," *IJARCCCE*, vol. 6, no. 6, pp. 11–14, Jun. 2017.
- [3] C. Patel, D. Shah, and A. Patel, "Automatic number plate recognition system (ANPR): A Survey," *Int. J. Comput. Appl.*, vol. 69, no. 9, pp. 21–33, 2013.
- [4] N. F. Gazcón, C. I. Chesñevar, and S. M. Castro, "Automatic vehicle identification for Argentinean license plates using intelligent template matching," *Pattern Recognit. Lett.*, vol. 33, no. 9, pp. 1066–1074, Jul. 2012.
- [5] R. P. . Persada, S. Aulia, B. D., and S. H., "Automatic face and VLP's recognition for smart parking system," *TELKOMNIKA (Telecommunication Comput. Electron. Control.)*, vol. 17, no. 4, pp. 1698, Aug. 2019.
- [6] S. Sugar, Lawade, S. Gayatri, Hedau, A. C, and Rangirwar, "Fake currency detection using image processing and other standard methods," *Int. J. Res. Sci. Eng.*, vol. 1, no. 1, pp. 128–131, 2014.
- [7] N. Sahu and M. Sonkusare, "A study on optical character recognition techniques," *Int. J. Comput. Sci. Inf. Technol. Control Eng.*, vol. 4, no. 1, pp. 01–15, 2017.
- [8] A. J. I. George and F. Gafoor, "Contourlet transform based feature extraction for handwritten Malayalam character recognition using neural network," *Int. J. Ind. Electron. Electr. Eng.*, vol. 2, no. 4, pp. 19–22, 2014.
- [9] G. Mathur and M. S. Rikhari, "Text detection in document images: highlight on using FAST algorithm," *Int. J. Adv. Eng. Res. Sci.*, vol. 4, no. 3, pp. 275–284, 2017.
- [10] M. Sonkusare and N. Sahu, "A Survey on handwritten character recognition (HCR) techniques for English alphabets," *Adv. Vis. Comput. An Int. J.*, vol. 3, no. 1, pp. 1–12, Mar. 2016.
- [11] V. L. Sahu and B. Kubde, "Offline handwritten character recognition techniques using neural network : A review," *Int. J. Sci. Eng. Res.*, vol. 1, no. 1–3, pp. 87–94, 2013.
- [12] S. Singh, "Optical character recognition techniques: A survey," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 2, no. 6, pp. 2009–2015, 2013.
- [13] M. Anif *et al.*, "Pengembangan aplikasi text recognition dengan klasifikasi neural network pada huruf Hijaiyah Gundul," *BIT*, vol. 10, no. 1, pp. 59–67, 2013.
- [14] S. Das and S. Banerjee, "An Algorithm for Japanese character recognition," *Int. J. Image, Graph. Signal Process.*, vol. 7, no. 1, pp. 9–15, Dec. 2014. .
- [15] S. R. Zanwar, U. B. Shinde, A. S. Narote, and S. P. Narote, "A comprehensive survey on soft computing based optical character recognition techniques," *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 978–987, 2019.
- [16] K. Dholakia, "A Survey on handwritten character recognition techniques for various Indian languages," *Int. J. Comput. Appl.*, vol. 115, no. 1, pp. 17–21, 2015.
- [17] U. Pal, R. Jayadevan, and N. Sharma, "Handwriting recognition in indian regional scripts: A survey of offline techniques," *ACM Trans. Asian Lang. Inf. Process.*, vol. 11, no. 1, 2012.
- [18] S. Impedovo, L. Ottaviano, and S. Occhinegro, "Optical Character Recognition — A Survey," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 4, no. 6, pp. 545–550, Jun. 2013.
- [19] L. Onyejebu and O. Ikechukwu, "Optical character recognition as a cloud service in Azure architecture," *Int. J. Comput. Appl.*, vol. 146, no. 13, pp. 14–20, 2016.
- [20] M. Luthfi, S. A. S. Agung, B. Prasetijo, and M. Somantri, "Pendeteksi sinyal jual/beli saham dengan fuzzy rule-based evidential reasoning dan C-means clustering," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 7, no. 3, Sep. 2018.
- [21] I. A. Doush and A. M. Al Trad, "Improving post-processing optical character recognition documents with Arabic language using spelling error detection and correction," *Int. J. Reason. Intell. Syst.*, vol. 8, no. 3/4, pp. 91, 2016.
- [22] R. Siregar, "Implementasi OTSU thresholding pada optical character recognition menggunakan engine tesseract," *J. Ilm. Core It*, vol. 7, no. 1, pp. 27–34, 2019.
- [23] G. Andersson and J. Isaksen, "Best practices in capacity building in public finance management in Africa: Experiences of NORAD and SIDA," *Rep. - Chr. Michelsen Inst.*, vol. 8, no. 16, pp. 1–53, 2002.
- [24] R. R. Chakravarthi and E. Chandra, "Digital signature representation using SOM method in OCR technique," *Int. J. Latest Eng. Manag. Res.*, vol. 02, no. 08, pp. 1–11, 2017.
- [25] C. Patel, A. Patel, and D. Patel, "Optical character recognition by open source OCR tool tesseract: A case study," *Int. J. Comput. Appl.*, vol. 55, no. 10, pp. 50–56, 2012.
- [26] C. Pratama, S. Aulia, D. N. Ramadan, and S. Hadiyoso, "Vehicle license plate detection for parking offenders using automatic license-plate recognition," *J. Southwest Jiaotong Univ.*, vol. 55, no. 4, 2020.
- [27] M. U. Niam, S. Aulia, B. Dirgantoro, and S. Hadiyoso, "System development for sale and buy execution in stock trading using knowledge-based system," *Int. J. Emerg. Technol.*, vol. 12, no. 1, pp. 6, 2021.