

Fraud Analytics: Solusi Preventif Terhadap Penyalahgunaan Promo

Studi Kasus: BLANJA.com (PT. MetraPlasa)

Iqbal Hanif/ Nik. 940088

Data Scientist/ Divisi Digital Service

PT. Telkom Indonesia

940088@telkom.co.id dan no. Hp: 085268207668

Ryan Andhika Perdana/ Nik. 876884

Data Scientist/ Divisi Digital Service

PT. Telkom Indonesia

876884@telkom.co.id dan no. Hp: 082246592232

Abstrak—BLANJA.com sebagai salah satu anak perusahaan Telkom Indonesia yang bergerak di segmen *e-commerce* mampu menempati posisi 10 besar *e-commerce* di Indonesia, namun BLANJA.com belum mencapai performa optimal dari sisi pendapatan. Hasil identifikasi masalah menunjukkan bahwa proses pengecekan transaksi secara manual dengan sumber daya terbatas menyebabkan *approval* transaksi di BLANJA.com menjadi lambat dan terdapat beberapa *fraud* yang lolos dari pengecekan transaksi. Oleh sebab itu, perlu ditemukan solusi untuk menganalisa dan mencegah transaksi *fraud* serta sistem otomasi agar pengecekan transaksi tidak berlangsung lama; solusi ini dikenal dengan istilah *fraud analytics*. Proses pembangunan *fraud analytics* melibatkan beberapa metode, diantaranya metode *substring*, *regular expression*, *fuzzy string matching*, *user behavior tracking*, serta otomasi melalui skema *Extract, Transform, Load (ETL)*.

Kata Kunci—*analytics, e-commerce, fraud*

1. LATAR BELAKANG

Indonesia, sebagai negara dengan jumlah penduduk terbesar keempat di dunia^[1], ternyata memiliki pertumbuhan jumlah pengguna internet yang cukup signifikan. Berdasarkan hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), total pengguna jasa internet di Indonesia tahun 2017 adalah 143,26 juta pengguna, bertambah sebesar 10,56 juta bila dibandingkan dengan tahun sebelumnya (2016 = 132,7 juta) dan 30,06 juta bila dibandingkan dengan dua tahun sebelumnya (2015 = 110,2 juta); informasi lebih detail tentang ini bisa dilihat di Gbr. 1^[2]. Persentase pengguna internet terhadap populasi Indonesia juga tumbuh dari 51.8% di tahun 2016 (dari total populasi 256,2 juta jiwa) menjadi 54.68% di tahun 2017 (dari total populasi 262 juta jiwa)^[3]. Tak pelak jika pertumbuhan pengguna jasa internet yang meliputi separuh populasi penduduk Indonesia ini memberikan dampak berupa perubahan perilaku dan gaya hidup masyarakat Indonesia, salah satunya adalah perilaku dan gaya hidup di bidang ekonomi. Masih dari hasil survei yang sama, 45.14% dari pengguna internet di Indonesia melakukan pengecekan atau pencarian harga via internet, 32,19% melakukan pembelian barang via internet, 17,04% melakukan transaksi perbankan via internet, bahkan 16,63% menjadi penjual dan menjual barang via internet^[4].

Meningkatnya penetrasi pengguna jasa internet tidak lepas dari pertumbuhan penggunaan *smartphone* di Indonesia. Menurut laporan *Indonesia Digital Landscape* dari GDP Ventures, 90% pengguna internet di Indonesia mengakses internet melalui *smartphone*. Fakta lain menunjukkan bahwa 88% pengguna *smartphone* di Indonesia lebih memilih menggunakan *mobile apps* yang terkoneksi dengan internet dibandingkan menggunakan *web browser*^[5]; hal ini menjadi salah satu faktor yang mendorong berkembang-pesatnya *start up* yang mengembangkan *mobile apps* di Indonesia, termasuk di segmen jual beli *online* atau yang lebih dikenal dengan istilah *e-commerce*. Masih dari laporan yang sama, kemudahan transaksi melalui *mobile apps* meningkatkan jumlah pengguna *e-commerce* hingga 35 juta lebih pengguna di tahun 2017, dengan rata-rata waktu akses *shopping apps* selama 90 menit

[1] "World Population Data Sheet 2017", Population Reference Bureau (PRB), 2018

[2] "Infografis Penetrasi & Perilaku Pengguna Internet Indonesia: Survey 2017", Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), 2018.

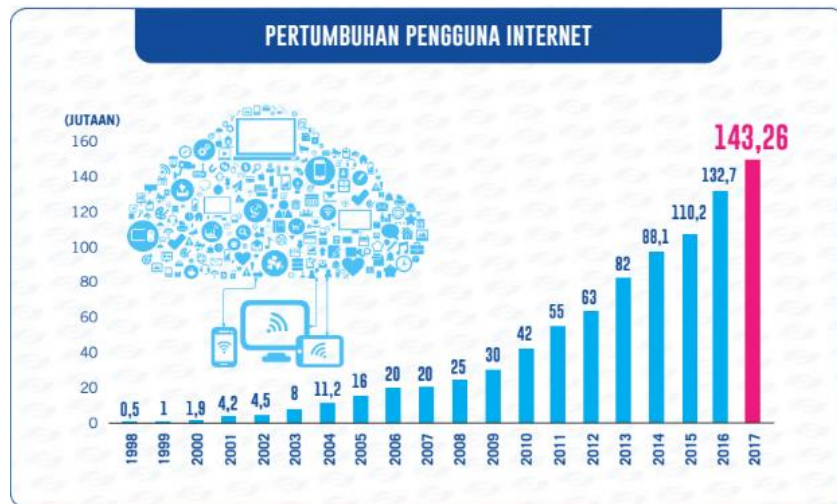
[3] "Infografis Penetrasi & Perilaku Pengguna Internet Indonesia: Survey 2016", Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), 2017.

[4] Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), *Op.Cit.*, 2018.

[5] "Indonesia Digital Landscape 2018", GDP Venture, 2018.

per bulan mejadikan Indonesia sebagai negara nomor dua di dunia dengan waktu rata-rata akses *shopping apps* terlama^[6].

Gbr. 1: Pertumbuhan Pengguna Internet di Indonesia (Sumber: Survei APJII)



Melihat tren tersebut, Telkom Indonesia sebagai perusahaan berbasis digital terbesar di Indonesia tentu tidak melewatkan peluang bisnis di bidang *online shopping* atau *e-commerce*. Telkom Indonesia bersama e-Bay mendirikan BLANJA.com (PT. MetraPlasa) pada tanggal 8 Desember 2014 melalui skema *joint venture*. Tiga tahun berselang, BLANJA.com masih mampu bersaing dalam 10 besar *e-commerce* Indonesia. Berdasarkan rilis situs aseanup.com, BLANJA.com menempati posisi ketujuh di Indonesia dengan perkiraan trafik per bulan sejumlah 11,69 juta^[7]. Rilis dari situs digitalnewsasia.com juga menempatkan BLANJA.com dalam 10 besar *e-commerce* Indonesia, dimana BLANJA.com menempati urutan sembilan di triwulan-1, naik ke urutan delapan di triwulan-2, naik ke urutan enam di triwulan-3, dan turun ke urutan tujuh di triwulan-4; data lebih detail tentang ini bisa dilihat di Gbr. 2^[8].

Gbr. 2: Top 10 Most Visited E-Commerce (Sumber: digitalnewsasia.com)



[6] Ibid.

[7] Anonim. "Top 10 e-commerce sites in Indonesia 2017", (<https://aseanup.com/top-e-commerce-sites-indonesia>, diakses 11 Maret, 2018)

[8] Anonim. "2017 e-commerce review for Indonesia", (<https://www.digitalnewsasia.com/digital-economy/2017-e-commerce-review-indonesia>, diakses 11 Maret, 2018)

Untuk memaksimalkan performansi BLANJA.com terutama dari segi pencapaian pendapatan, Big Data DDS memberikan dukungan kepada BLANJA.com. Tim penulis yang berasal dari kelompok Data Scientist DDS secara intensif membangun komunikasi dengan pihak BLANJA.com untuk memberikan dukungan dan solusi terhadap permasalahan yang ada di BLANJA.com. Sesuai permintaan dari pihak BLANJA.com, salah satu solusi yang kami tawarkan saat ini adalah *fraud analytics*, sebagai jalan keluar pencegahan penyalahgunaan promo.

2. IDENTIFIKASI PERMASALAHAN

A. Sumber Informasi

Sumber utama informasi yang digunakan untuk mengidentifikasi masalah adalah forum diskusi yang dilakukan dengan dua unit di BLANJA.com: unit Business Analyst (BA) dan unit Trust & Security (T&S). Forum diskusi dilakukan baik secara *online* maupun *offline*. Forum diskusi *offline* diadakan setiap minggu yang berlokasi di kantor BLANJA.com. Dalam diskusi tersebut, progress pengembangan *fraud analytics* akan direview dan dievaluasi untuk perbaikan dan pengembangan di tahap berikutnya. Semua hasil diskusi tercatat dalam bentuk *Minutes of Meeting* (MoM) dan dibagikan kepada kedua belah pihak (DDS dan BLANJA.com) serta dilaporkan ke atasan setiap minggunya dalam bentuk *highlight activities report* (eviden diskusi dan pelaporan *progress* bisa dilihat di Gbr. 4). Forum diskusi *online* menggunakan aplikasi pesan WhatsApp, dimana tim penulis berada satu grup dengan tim dari BLANJA.com sehingga bisa dengan bebas membahas hal-hal yang terkait dengan sistem *fraud detection* yang berjalan saat ini di BLANJA.com.

B. Fraud di BLANJA.com

Ada dua tipe *fraud* yang teridentifikasi oleh tim penulis, yaitu transaksi yang dilakukan oleh trader dan metode pembayaran yang tercatat berbeda dengan metode pembayaran yang disyaratkan.

Gbr. 3: Diskusi dengan Tim T&S dari BLANJA.com dan Progress Report dengan Atasan Unit Data Scientist DDS



Diskusi dengan tim T&S



Progress report dengan atasan

1) Transaksi yang dilakukan *trader*

Trader adalah penjual yang berperan sebagai pembeli fiktif di BLANJA.com, dimana mereka membuat pembelian fiktif dengan menggunakan promo agar memperoleh keuntungan. Ada beberapa trik yang digunakan *trader* untuk memperoleh banyak kupon promo, diantaranya membuat banyak *user name* dengan membuat *email* yang mirip atau rekursif dengan penambahan angka atau karakter khusus dalam jumlah yang cukup banyak. Contohnya adalah menambahkan titik diantara alamat *email* dengan posisi yang berpindah-pindah, dimana nantinya pesan yang masuk dapat diatur agar hanya masuk ke satu kotak masuk di satu *email* dengan melakukan pengaturan *email*.

2) Metode pembayaran yang tercatat berbeda dengan metode pembayaran yang disyaratkan

BLANJA.com cukup sering mengadakan promo baik promo langsung dari BLANJA.com maupun promo melalui kerja sama dengan mitra yang lazimnya memiliki beberapa syarat dalam transaksi dengan harapan mampu meningkatkan jumlah pengguna dan jumlah transaksi. Ada beberapa temuan terkait transaksi promo yang tidak memenuhi syarat promo, sehingga promo menjadi tidak tepat sasaran. Hal ini dikarenakan adanya masalah pencatatan metode pembayaran dalam sistem BLANJA.com, dimana metode pembayaran yang tercatat dalam sistem laporan tidak sesuai dengan metode pembayaran sebenarnya.

3. PENYELESAIAN PERMASALAHAN

Metode penyelesaian masalah dibagi dalam tiga lingkup utama mengikuti hasil analisa masalah: verifikasi metode pembayaran, identifikasi *trader*, *user behavior tracking*, dan otomasi proses deteksi *fraud*. Semua metode diterapkan melalui dua bahasa pemrograman yaitu bahasa *query* atau lebih dikenal dengan istilah *Structured Query Language* (SQL) dan bahasa pemrograman Python. Proses otomasi dilakukan dengan menggunakan tools *Extract, Transform, Load* (ETL) yaitu Talend.

A. Verifikasi Metode Pembayaran

Metode pembayaran riil yang digunakan pembeli di BLANJA.com dapat diperoleh di tabel “t_trade_payment_notification_log”, yaitu tabel yang berisikan data yang bersumber dari *online payment gateway* dimana data transaksi tersimpan dalam bentuk *string* yang panjang dan menyimpan semua informasi transaksi mulai dari metode pembayaran, status pembayaran, dan sebagainya. Untuk mengambil informasi di *string* tersebut, digunakan metode *substring*, dimana hasil dari *substring* tersebut menjadi metadata untuk memperoleh informasi terkait metode pembayaran^[11]. Proses *substring* dilakukan dengan menggunakan bahasa SQL, dengan rincian rumus sebagai berikut:

SUBSTRING (*string*, *start*, *length*) (1)

dengan *string*/teks yang akan diolah dengan metode *substring*, *start* adalah lokasi karakter pertama yang akan diambil, dan *length* adalah jumlah atau panjang karakter yang akan diambil.

Dengan metode *substring*, *string* yang tersimpan di tabel “t_trade_payment_notification_log” dibagi menjadi beberapa bagian, kemudian diambil bagian penting yang mengandung informasi yang dibutuhkan (detail dari bagian-bagian tersebut dapat dilihat di Tabel 1).

TABEL 1: DESKRIPSI META-DATA HASIL METODE SUBSTRING

Nama Kolom	Deskripsi
PAYMENT_TYPE_LOG	Tipe pembayaran yang digunakan pembeli
BANK	Nama bank yang digunakan pembeli untuk melakukan transaksi pembelian
BIN_NUMBER	Bank Identification Number
FRAUD_STATUS	Status <i>Fraud</i> yang dikeluarkan oleh sistem, jika di approve maka transaksi akan masuk ke tahap selanjutnya (diteruskan ke penjual)
SITE_ID	Jenis promo (bekerja sama dengan mitra) yang digunakan oleh pembeli

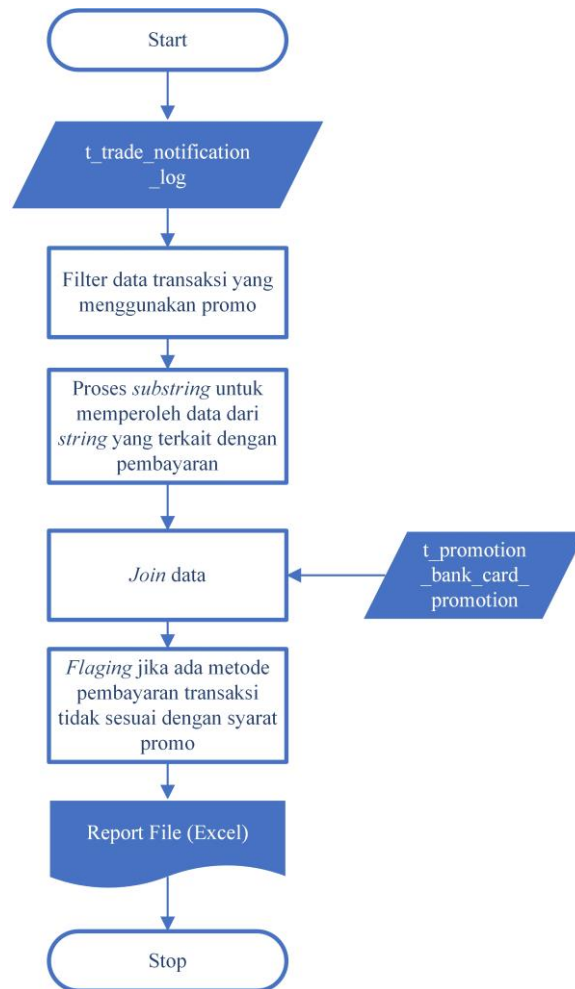
[6] Ibid.

[7] Anonim. "Top 10 e-commerce sites in Indonesia 2017", (<https://aseanup.com/top-e-commerce-sites-indonesia>, diakses 11 Maret, 2018)

[8] Anonim. "2017 e-commerce review for Indonesia", (<https://www.digitalnewsasia.com/digital-economy/2017-e-commerce-review-indonesia>, diakses 11 Maret, 2018)

Informasi yang diperoleh dari “t_trade_payment_notification_log” akan disandingkan dengan data promo di BLANJA.com yang ada di tabel “t_promotion_bank_card”, dan akan dilakukan pengecekan apakah metode pembayaran riil yang dilakukan oleh pembeli di BLANJA.com saat menggunakan kupon promo sesuai dengan metode pembayaran yang disyaratkan oleh promo tersebut. Proses kerja yang dilakukan untuk melakukan pengecekan tersebut disajikan pada diagram alur di bawah ini.

Gbr. 4: Diagram Alur Proses Verifikasi Metode Pembayaran



Proses *flagging* dilakukan ketika metode pembayaran sebenarnya tidak sesuai dengan metode pembayaran yang disyaratkan dengan metode pembayaran yang disyaratkan promo. Hasil laporan verifikasi metode pembayaran ini kemudian dikirim ke tim T&S untuk selanjutnya dikirimkan ke *admin* BLANJA.com untuk mencegah transaksi yang dilakukan oleh *trader* tersebut.

B. Identifikasi Trader

Identifikasi *trader* dilakukan secara harian untuk memeriksa apakah transaksi pada hari tersebut terdapat transaksi yang berasal dari *trader* yang ada di dalam *blacklist*. Selain menggunakan metode pencocokan biasa, metode pengecekan harian juga dikembangkan dengan menggunakan metode penghapusan *special & numeric character* menggunakan *reguler expression* dan metode *fuzzy string matching*.

1) Penghapusan Special & Numeric Character

Penghapusan *special character* dilakukan khusus untuk pemeriksaan *email* pengguna atau pembeli di BLANJA.com, mengingat ada beberapa trik yang dapat digunakan oleh *trader* untuk menyiasati sistem pemberian kupon BLANJA.com (contoh: pemberian kupon untuk satu *user name* atau *email* dapat disiasati dengan membuat *email* yang mirip dengan menyertakan karakter titik di tengahnya). Metode ini dilakukan

karena lebih sederhana dibanding metode *fuzzy string matching* dan proses komputasi yang lebih ringan sehingga lebih hemat waktu dan sumber daya. Dalam metode ini, digunakan metode *regular expression*, yaitu metode pengolahan teks yang kompleks secara otomatis dengan menggunakan notasi *pattern* yang terdiri atas meta-karakter sebagai penanda *pattern* dan repetisi sebagai penanda pengulangan *pattern* ^[12]. Daftar meta-karakter dalam *regular expression* ditunjukkan pada Tabel 2.

TABEL 2: DAFTAR META-KARAKTER DALAM REGULAR EXPRESSION

Simbol	Fungsi
/	mengawali dan mengakhiri (mengapit) <i>pattern</i>
^	mencocokkan <i>pattern</i> yang terletak pada awal subjek
\$	mencocokkan <i>pattern</i> yang terletak pada akhir subjek
.	mencocokkan dengan karakter apapun, kecuali baris baru
.*?	mencocokkan dengan karakter apapun termasuk baris baru
[]	membuka dan menutup definisi <i>character class</i>
	tanda pemisah dari untuk opsi alternative
()	membuka dan menutup sub- <i>pattern</i>
\	karakter <i>escape</i>
{x, y}	pembilang repetisi dengan nilai minimal x dan maksimal y
?	pembilang repetisi minimal nol dan maksimal satu {0, }
*	pembilang repetisi minimal nol dan maksimal tidak terbatas {0, }
+	pembilang repetisi minimal satu dan maksimal tidak terbatas {1, }

Metode ini bisa diterapkan di Python melalui fungsi sebagai berikut:

$$df['column_name'].replace('[^A-Za-z\s]+', "", regex=True) \quad (3)$$

dengan *df* adalah tabel data (*data frame*), *column_name* ada nama kolom, dan `'[^A-Za-z\s]+'` adalah *pattern* yang digunakan untuk menjaga karakter alfabet dan menghapus karakter selain alfabet.

Hasil dari metode ini adalah *string* yang memuat *email* pengguna yang termodifikasi (tanpa angka dan *special character*). Pemanfaatan metode *regular expression* dalam proses identifikasi *trader* baru digambarkan pada diagram alur pada Gbr. 7. Proses *flagging* dilakukan ketika ada pengguna atau pembeli yang berbelanja di BLANJA.com terdapat dalam *blacklist trader* (hasil proses *regular expression*). Laporan hasil identifikasi *trader* harian ini kemudian dikirim ke tim T&S untuk selanjutnya dikirimkan ke *admin* BLANJA.com untuk mencegah transaksi yang dilakukan oleh *trader*.

2) Fuzzy String Matching

Metode *fuzzy string matching* merupakan salah satu metode pencarian *string* yang menggunakan proses pendekatan terhadap pola dari *string* yang dicari. Kunci dari konsep ini adalah bagaimana memutuskan bahwa sebuah *string* yang dicari memiliki kesamaan dengan *string* yang tertampung, meskipun tidak sama persis dalam susunan karakternya ^[13].

Ukuran kuantitatif yang digunakan untuk mengukur kemiripan antar suatu *string* dengan *string* yang lain dihitung dengan algoritma *Levenshtein Distance* atau sering disebut juga algoritma *Edit Distance*. Algoritma ini menghitung jumlah operasi penghapusan, penyisipan, dan penukaran yang harus dilakukan terhadap suatu *string* agar sama dengan *string* lain yang menjadi pembanding. Sebagai contoh, *string* “komputer” dan “computer” memiliki *distance* 1 karena hanya perlu dilakukan satu operasi saja untuk mengubah satu *string* ke *string* yang lain. Dalam kasus dua *string* di atas, *string* “computer” dapat menjadi “komputer” hanya dengan melakukan satu penukaran karakter “c” menjadi “k” ^[14].

Perhitungan skor kemiripan antar dua *string* *i* dan *j* dilakukan dengan melakukan operasi perhitungan seperti dalam rumus berikut:

[11] Aji T. M., Riyanto D. E., Wibawa H. A. "Penerapan Web Services dan Regular Expression untuk Verifikasi Alamat Berbasis Hasil Penelusuran Situs Kodepos PT Pos Indonesia". Journal of Informatics and Technology, Vol 1, No 1, Tahun 2012. 38-51

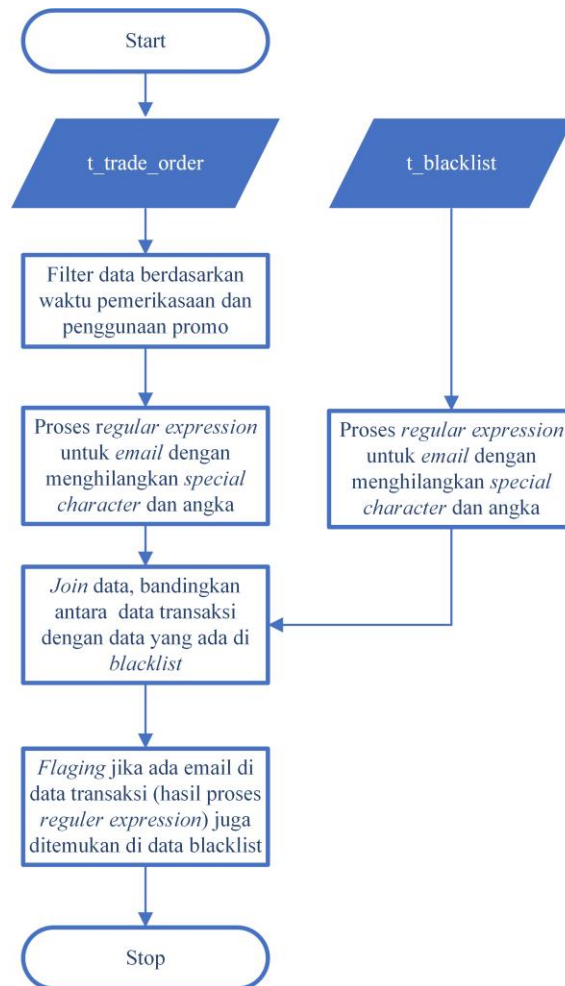
[12] Gurning A. I. A., Zarnelly, Adawiyah A. , "Penerapan Fuzzy String Matching pada Aplikasi Pencarian Tugas Akhir Mahasiswa Jurusan Sistem Informasi Berbasis Web (Studi Kasus: Fakultas Sains dan Teknologi UIN Suska Riau)". Jurnal Rekayasa dan Manajemen Sistem Informasi, Vol. 2, No. 1, Februari 2016, 54- 59

[13] Ibid.

$$\left(1 - \frac{p}{\max(i,j)}\right) \times 100 \quad (2)$$

dengan p adalah jumlah operasi yang dilakukan agar *string* i sama dengan j , dan $\max(i,j)$ adalah ukuran maksimum antara *string* i dan j , dimana *string* terpanjang menjadi patokan ukuran perhitungan skor kemiripan^[15]. Skor kemiripan tertinggi adalah 100 dan skor kemiripan terendah adalah 0.

Gbr. 5: Diagram Alur Proses Pemeriksaan harian Menggunakan Metode Regular Expression



Perhitungan skor ini cukup mudah dilakukan melalui fungsi Python berikut:

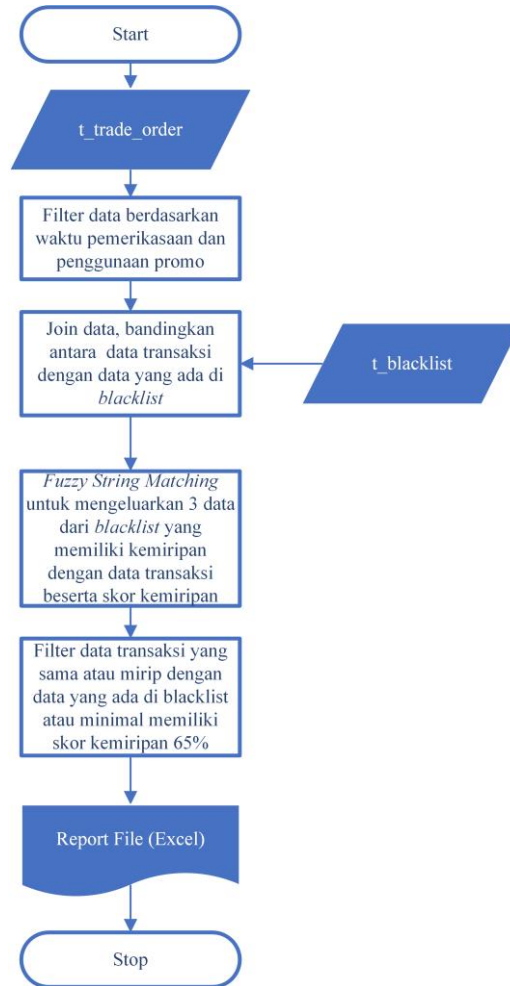
$$\text{fuzz.ratio}(\text{string_1}, \text{string_2}) \quad (3)$$

dengan *string_1* adalah *string* pertama dan *string_2* adalah *string* kedua yang dibandingkan dengan *string* pertama. Pemanfaatan metode *fuzzy string matching* dalam proses identifikasi *trader* harian digambarkan pada diagram alur pada Gbr. 8.

Proses *flagging* dilakukan ketika ada pengguna atau pembeli yang berbelanja di BLANJA.com terdapat dalam *blacklist trader*, atau memiliki kemiripan baik dari segi *user name*, *email*, kontak, maupun alamat penerima yang ada di *blacklist trader*. Laporan hasil identifikasi *trader* harian ini kemudian dikirim ke tim T&S untuk selanjutnya dikirimkan ke *admin* BLANJA.com untuk mencegah transaksi yang dilakukan oleh *trader* tersebut.

[14] Umer A. "C#: Calculating Percentage Similarity of 2 strings", (<https://social.technet.microsoft.com/wiki/contents/articles/26805-c-calculating-percentage-similarity-of-2-strings.aspx>, diakses pada 24 Maret, 2018)

Gbr. 6: Diagram Alur Pemeriksaan Harian Menggunakan Metode Fuzzy String Matching



C. User Behavior Tracking

User behavior tracking dengan memanfaatkan *browser cookies* merupakan salah satu metode mutakhir yang lazim digunakan oleh penyedia jasa *online analytics* dan *online advertisement* seperti Google dan Facebook. Dengan memasang *script* khusus ke dalam situs, *browser* akan secara otomatis mengirimkan data *cookies* berupa data aktivitas pengunjung situs, sehingga aktivitas pengguna dapat di telusuri dan disimpan datanya ke dalam *database*. Untuk situs BLANJA.com, data tersimpan dalam *lag* waktu satu hari, dimana data aktivitas pengguna hari ini baru bisa diakses keesokan harinya. Detil data yang tersimpan melalui metode ini dapat dilihat pada Tabel 3.

TABEL 3. DESKRIPSI TABEL BROWSER COOKIES DI DATABASE BIG DATA DDS

Nama Kolom	Deskripsi
date_id	ID Tanggal
Created	Tanggal terbentuknya <i>data cookies</i>
id_cookies	ID <i>Cookies</i>
id_session	ID <i>Session</i>
total_cookies	Jumlah <i>data cookies</i> daam satu <i>session</i>
buyer_email	Email pembeli

Konsep utama dalam penelusuran aktivitas pengguna situs BLANJA.com ini adalah *session id* dan *cookies id*. *Session id* akan otomatis tersimpan ketika pengguna situs melakukan aktivitas di dalam web dan tidak berhenti (*idle*) lebih dari 30 detik. Ketika pengguna web berhenti beraktivitas selama lebih dari 30 detik maka aktivitas selanjutnya

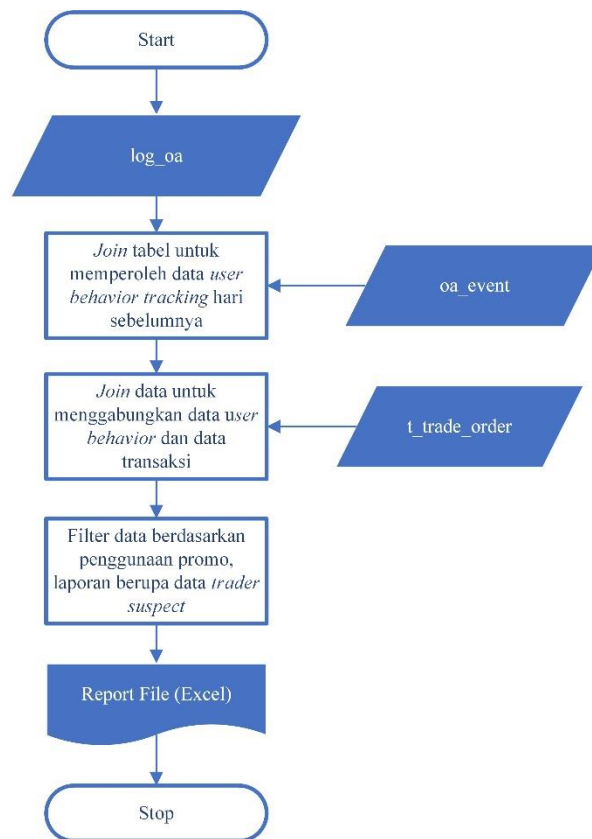
akan terekam di *session id* yang baru. Sedangkan *cookies id* adalah identitas dari browser dan perangkat yang digunakan ketika mengakses web BLANJA.com, yang akan tetap sama meskipun *session*-nya sudah habis atau pengguna berganti.

Pemanfaatan data *user behavior* untuk identifikasi *trader* dapat dilakukan dengan melakukan pengecekan berikut:

- Seberapa sering pengguna melakukan transaksi dalam satu *session*.
- Seberapa sering aktivitas transaksi di BLANJA.com dilakukan di suatu *browser* atau perangkat meskipun berganti pengguna dalam selang waktu tertentu.

Pemanfaatan metode *user behavior tracking* dalam proses identifikasi *trader* baru digambarkan pada diagram alur pada Gbr 9.

Gbr. 9: Diagram Alur Identifikasi Trader Menggunakan Data User Behavior Tracking



D. Otomasi Fraud Detection

Otomasi proses deteksi *fraud* dilakukan dengan menggunakan *Extraction, Transformation, dan Load* (ETL). ETL adalah salah satu proses pada *datawarehouse*, yaitu proses mengumpulkan data dari berbagai macam sumber dan mengolah data menjadi data yang bersih sesuai dengan ketentuan *datawarehouse*. Proses ETL pada umumnya terdiri dari berbagai macam aktivitas dan membutuhkan waktu serta memori yang cukup besar ^[16]. Proses ETL dengan menggunakan Talend secara keseluruhan dapat dilihat di Lampiran 1.

Proses ETL dalam *fraud detection* secara sederhana dapat dibagi ke dalam tiga tahap: *input*, *proses*, dan *output*.

1) Input

Tahap input merupakan tahap memasukkan data dari sumber data ke proses ETL. Dalam pengembangan *fraud analytics* BLANJA.com, sumber data yang digunakan adalah *database* transaksi (MySQL) milik BLANJA.com

2) Proses

Tahap proses adalah tahap pengolahan data mulai dari data mentah hingga menjadi informasi yang ingin disajikan dan dikeluarkan menjadi output ETL. ETL melibatkan beberapa proses yang bisa dilihat dalam Tabel 4. Contoh proses ETL dapat dilihat pada Gbr.10.

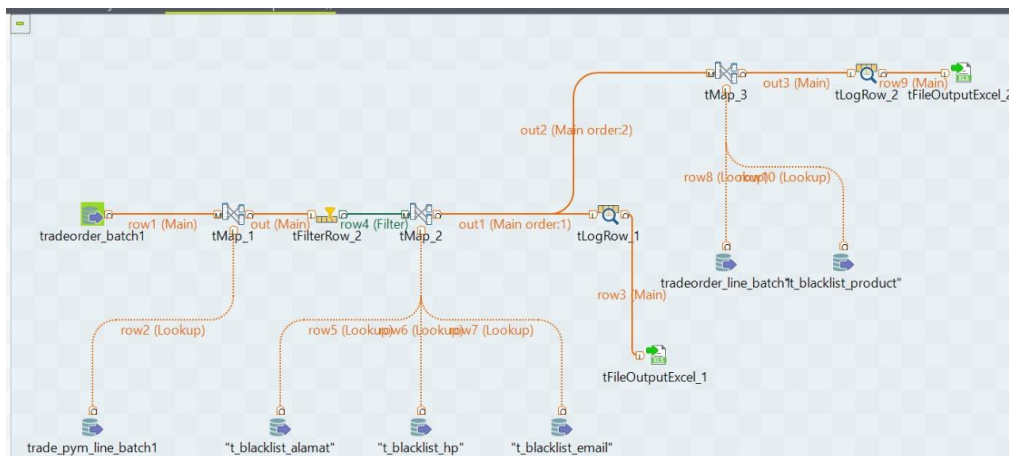
TABEL 4: DAFTAR PROSES YANG DILAKUKAN DI ETL (TALEND)

Nama Proses	Deskripsi
tInput	Proses memasukkan data dari sumber data ke dalam proses ETL
tMap	Proses <i>join</i> antara satu sumber data dengan sumber data yang lainnya
tSystem	Proses menggunakan bahasa pemrograman eksternal (contoh: Python)
tFilterRow	Proses filter data sesuai dengan kriteria tertentu
tLogRow	Proses menghasilkan data dalam bentuk log/text (untuk mengecek data sebelum dijadikan <i>file</i>)
tFileOutput	Proses menghasilkan output berupa file

3) Output

Tahap *output* merupakan tahap mengeluarkan data hasil proses menjadi tabulasi data yang disajikan dalam bentuk *file*. Dalam pengembangan *fraud analytics* BLANJA.com, *output file* dikeluarkan dalam bentuk *file* Microsoft Excel (.xlsx).

Gbr. 10: Contoh Proses ETL



E. Pemaparan Hasil Uji Coba

Berikut beberapa hasil uji coba *fraud analytics*:

1) Verifikasi Metode Pembayaran

Uji coba proses verifikasi metode pembayaran menggunakan data transaksi pada tanggal 13 Maret 2018 *batch* 2. Melalui proses ETL, proses verifikasi mampu dilakukan dalam waktu kurang lebih satu menit, dimana proses dijalankan pada pukul 09:52 dan selesai pada pukul 09:53 (proses ETL dan waktu proses dapat dilihat pada Gbr. 11). Hasil proses verifikasi dapat dilihat pada Lampiran 2.

2) Identifikasi Trader Harian

Uji coba identifikasi *trader* menggunakan data transaksi tanggal 27 Maret 2018. Proses identifikasi menggunakan bahasa Python dan dikerjakan melalui Jupyter Notebook (belum diintegrasikan ke ETL). Untuk tahap ini, kami mengeluarkan contoh berupa 40 data hasil identifikasi transaksi yang dicurigai sebagai transaksi *trader*. Untuk tahap ini, masih diperlukan pemeriksaan lebih lanjut menggunakan data kontak, alamat, dan *internet protocol* secara manual oleh tim T&S. Contoh data identifikasi *trader* harian bisa dilihat di Lampiran 3.

3) User Behavior Tracking

Uji coba pengambilan data *user behavior tracking* menggunakan data *browser cookies* tanggal 14 Maret 2018. Melalui proses ETL, proses verifikasi mampu dilakukan dalam waktu 89.99 detik atau kurang lebih satu setengah menit (proses ETL dan waktu proses dapat dilihat pada Gbr. 12). Data *user behavior tracking* dapat dilihat di Lampiran 4.

4. KESIMPULAN & SARAN

A. Kesimpulan

Fraud analytics yang dikembangkan mampu memangkas waktu verifikasi metode pembayaran dan identifikasi harian dibandingkan dengan metode manual yang saat ini berlangsung di BLANJA.com. Beberapa *suspect transaction* dapat diidentifikasi sehingga kedepannya diharapkan mampu menyelamatkan biaya *marketing* BLANJA.com yang digunakan untuk mengadakan promo. Data *user behavior tracking* juga menjadi alternatif yang bermanfaat bagi BLANJA.com mengingat saat ini *trader blacklist update* masih dilakukan dengan pemeriksaan manual dan dilakukan perminggu. Namun, verifikasi dan keputusan pembatalan transaksi tetap harus melalui keputusan dari tim T&S.

B. Saran & Next Plan

Hasil ujicoba dirasa masih belum optimal dikarenakan terbatasnya sumber daya untuk melakukan komputasi yang cukup berat (spesifikasi komputer kurang memadai, hanya ada satu komputer yang spesifikasinya memadai dan hanya boleh diakses di kantor BLANJA.com) serta akses data transaksi yang terbatas (hanya boleh diakses satu user). Kedepannya akan diajukan perangkat komputer yang lebih canggih dan akses data ke tim Data Scientist DDS lebih dari satu agar tim pengembang dapat berkolaborasi mengembangkan *fraud analytics* tanpa harus bergantian mengakses data BLANJA.com. Tim penulis juga akan mengembangkan *dashboard* untuk mempermudah tim BLANJA.com dalam mengakses data hasil proses *fraud analytics*, melakukan validasi metode *scoring* menggunakan metode *cross validation* dan *train-test* menggunakan data BLANJA yang lama, serta melakukan *crawling* baik data e-commerce kompetitor maupun *crawling* sosial media sebagai sumber data alternatif dalam identifikasi *trader*.

Daftar Pustaka

- [1] “World Population Data Sheet 2017”, Population Reference Bureau (PRB), 2018
- [2][4] “Infografis Penetrasi & Perilaku Pengguna Internet Indonesia: Survey 2017”, Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), 2018.
- [3] “Infografis Penetrasi & Perilaku Pengguna Internet Indonesia: Survey 2016”, Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), 2017.
- [5][6] “Indonesia Digital Landscape 2018”, GDP Venture, 2018.
- [7] Anonim. "Top 10 e-commerce sites in Indonesia 2017", (<https://aseanup.com/top-e-commerce-sites-indonesia>, diakses 11 Maret, 2018)
- [8] Anonim. "2017 e-commerce review for Indonesia", (<https://www.digitalnewsasia.com/digitaleconomy/2017ecommerce-review-indonesia>, diakses 11 Maret, 2018)
- [9] “Laporan Mingguan Performansi CFU DS Periode W1 Maret (1- 8 Mar. 2018)”, Media & Digital Business Telkom Indonesia, 2018.
- [10] “Laporan Performansi Transaksi, Trafik, dan Revenue BLANJA (Jan -Dec 2017)”, Digital Commerce & Logistic - Divisi Digital Service Telkom Indonesia, 2017.
- [11] Sugianto A., "Metode Identifikasi Permasalahan dan Kebutuhan Masyarakat" Peper Pemberdayaan Masyarakat Pedesaan, 2012.
- [12] Madsu Y. M. "Teknik Substring Weighting untuk Konversi Mata Kuliah". Jurnal Ilmiah Teknologi Informasi Terapan. Vol. 2 No. 2, April 2016. 157-161 [13]
- [13] Aji T. M., Riyanto D. E., Wibawa H. A. "Penerapan Web Services dan Regular Expression untuk Verifikasi Alamat Berbasis Hasil Penelusuran Situs Kodepos PT Pos Indonesia". Journal of Informatics and Technology, Vol 1, No 1, Tahun 2012. 38-51
- [14][15] Gurning A. I. A., Zarnelly, Adawiyah A. , "Penerapan Fuzzy String Matching pada Aplikasi Pencarian Tugas Akhir Mahasiswa Jurusan Sistem Informasi Berbasis Web (Studi Kasus: Fakultas Sains dan Teknologi UIN Suska Riau)". Jurnal Rekayasa dan Manajemen Sistem Informasi, Vol. 2, No. 1, Februari 2016, 54- 59
- [16] Ibid.Umer A. "C#: Calculating Percentage Similarity of 2 strings", (<https://social.technet.microsoft.com/wiki/contents/articles/26805.c-calculating-percentage-similarity-of-2-strings.aspx>, accessed on March 24, 2018)
- [17] Dharayani R., Laksitowening K. A, Yanuarfiani A. P. “Implementasi ETL (Extract, Transform, Load) Pangkalan Data Perguruan Tinggi dengan Menggunakan State-Space Problem”, e-Proceeding of Engineering, Vol.2, No.1 April 2015. 1159-1165