Sistem Rekomendasi Wedding Organizer Menggunakan Metode Content-Based Filtering Dengan Algoritma Random Forest Regression

Dewi Ayu Pratiwi[1], Anita Qoiriah[2]
[1,2] Jurusan Teknik Informatika, Universitas Negeri Surabaya
[1]dewi.17051204005@mhs.unesa.ac.id
[2]Anita Qoiriah@unesa.ac.id

***Abstrak*— Dewasa ini, jasa vendor pernikahan banyak diminati calon pengantin dalam mempersiapkan acara pernikahan agar efisiensi waktu dan biaya. Selalu ada problematika ketika memilih penyedia jasa pernikahan diantaranya adalah *budget* dari pengantin akan mendapat fasilitas apa saja dari konsep acara yang diinginkan serta wedding organizer mana yang akan dipilih yang sesuai dengan budget dan letaknya tidak jauh. Untuk memudahkan calon pengantin dalam menemukan *wedding organizer* yang sesuai maka diperlukan perhitungan khusus dengan kriteria tertentu dari calon pengantin. Maka dilakukan penelitian dengan membentuk sistem rekomendasi *wedding organizer* dalam mendukung pengambilan keputusan. Sistem rekomendasi merupakan aplikasi yang menyediakan serta merekomendasikan item ketika *user* menentukan keputusan. Sistem rekomendasi disini menggunakan metode *Content-Based Filtering* dan algoritma *Random Forest Regression*. *Content-Based Filtering* merupakan metode yang biasa digunakan untuk membangun sistem rekomendasi. *Content-based filtering* pada penelitian ini memberikan rekomendasi berdasarkan hasil analisis kemiripan item di beberapa atribut yang akan menghasilkan nilai kesamaan berupa *cosine similarity*. Kemudian *cosine similarity* akan diproses dengan metode *Random Forest Regression* sehingga menghasilkan model yang mampu menghitung nilai kesamaan dari dua paket *wedding organizer*. Nilai kesamaan yang tertinggi akan menjadi hasil rekomendasi. Digunakannya *Random Forest Regression* karena merupakan algoritma *machine learning* yang bagus dalam memecahkan masalah klasifikasi maupun regresi. Dari pengujian menggunakan pemodelan *Random Forest Regression* memeproleh hasil akurasi 83,750% dengan nilai MAPE sebesar 16,249% pada pengujian K=10 dalam 100 *tree random forest*. Penelitian dengan membentuk sistem rekomendasi *wedding organizer* ini diharapkan dapat memudahkan calon pengantin dalam memilih *wedding organizer* untuk kelancaran dalam menyiapkan pesta pernikahan.**

***Kata Kunci*— Wedding Organizer, Sistem Rekomendasi, Sistem Pendukung Keputusan, Content-Based Filtering, Random Forest.**

I. PENDAHULUAN

Wedding organizer merupakan penyedia jasa yang secara khusus membantu calon pengantin mempersiapkan segala sesuatu yang dibutuhkan untuk menggelar sebuah pesta pernikahan. Banyak dari calon pengantin dan keluarga yang memanfaatkan jasa *wedding organizer* dikarenakan kesibukan yang mereka miliki sehingga sedikit waktu yang dapat digunakan untuk mencari informasi terkait *wedding organizer* yang dipilih. Keberadaan *wedding organizer* sendiri merupakan suatu hal yang yang dibutuhkan bagi semua orang terutama calon pengantin. Dalam keadaan zaman yang sekarang semakin modern menjadikan banyak bermunculan penyedia jasa pernikahan sehingga seringkali calon pengantin merasa bingung untuk memilih penyedia jasa pernikahan tersebut. Kebanyakan yang dialami calon pengantin adalah sedikitnya waktu luang mereka untuk sekedar survey di tempat *wedding organizer.* Selain itu banyak dari mereka juga kurang informasi atau bisa juga kelebihan informasi terkait keberadaan *wedding organizer* yang berarti. mereka kurang tahu informasi dimana wedding organizer yang ada di dekat lingkungan mereka dan untuk mereka yang sudah mendapat banyak informasi yaitu mereka mengetahui mengenai keberadaan *wedding organizer* tapi tidak mengetahui fasilitas atau paket-paket yang tersedia dari wedding organizer tersebut.

Seiring dengan kemajuan teknologi yang pesat pada zaman modern saat ini menjadikan banyak perubahan pada kehidupan masyarakat yang mana membuat penetrasi teknologi digital semakin menaik. Masyarakat cenderung membutuhkan informasi yang cepat, akurat, dan terpercaya sehingga penggunaan media teknologi informasi serta pemanfaatan internet tidak bisa lepas dari upaya mereka dalam mencari informasi. Selain itu, manfaat perkembangan teknologi salah satunya adalah berbagi informasi menjadi lebih mudah seperti informasi penyedia jasa pernikahan untuk keperluan persiapan acara pernikahan karena sebagian besar pasangan menyewa jasa *wedding organizer* untuk membantu agar pernikahan mereka berjalan lancar. Informasi rekomendasi *wedding organizer* pilihan dari saudara atau rekan pasti banyak (23,6%) namun, juga masih banyak dijumpai mereka yang mencari informasi melalui internet seperti lewat *platform Bridestory* (16,4%), sosial media (13,2%), dan *search engine* sendiri (10,7%) merupakan preferensi andalan dalam menemukan vendor pernikahan [1]. Sementara itu, pilihan *wedding organizer* yang terdapat di internet juga sangat banyak dengan berbagai macam penawaran dan konsep yang menarik sehingga membuat para calon pengantin bingung dalam memilih jasa pernikahan tersebut.

Meninjau dari permasahan di atas disimpulkan bahwasanya masyarakat menginginkan untuk mendapat informasi dengan mudah, dimana letak *wedding organizer* yang dekat dengan tempat tinggal atau pesta pernikahan dilangsungkan serta perkiraan fasilitas yang didapatkan berdasarkan *budget* yang dianggarkan dari keluarga tersebut. Hal ini memunculkan sebuah gagasan untuk penulis membuat sebuah sistem berupa rekomendasi *wedding organizer* yang akan memberikan

rekomendasi bagi pengguna yang kesulitan dalam menentukan pilihan *wedding organizer* yang ingin digunakan.

Sistem rekomendasi dapat membantu pengguna dari masalah informasi yang berlebihan dan menemukan *item* yang mungkin diinginkan. Sistem rekomendasi merupakan suatu aplikasi untuk menyediakan dan merekomendasikan *item* dalam membuat keputusan yang diinginkan pengguna [2]. Pengertian lain dari sistem rekomendasi yaitu merupakan suatu alat dan teknik dalam sebuah *software* yang bisa memberikan saran kepada pengguna terkait *item* yang dapat bermanfaat bagi pengguna dalam menentukan pilihannya [3]. Sistem rekomendasi berpotensi memprediksi sejumlah *item* untuk pengguna. Jumlah informasi yang tersedia sekarang terus meningkat sehingga menimbulkan masalah yaitu pengguna kesulitan memilih *item* yang sesuai dengan kebutuhan, sehingga masuklah solusi berupa sistem rekomendasi. Guna membangun sistem rekomendasi, diperlukan metode dalam menyelesaikan permasalahan, salah satunya metode *Content-Based Filtering* yang akan diterapkan penulis dalam penelitian ini.

Meode *Content-Based Filtering* merupakan sistem yang merekomendasikan sesuatu yang cocok untuk pengguna berdasarkan deskripsi *item* dan preferensi pengguna. *Content-Based Filtering* menggunakan informasi dari *item* yang dijadikan atribut sebagai proses dalam menghitung kemiripan antar *item*. Secara umum, metode *Content-Based Filtering* membentuk profil pengguna berdasar atribut pembentuk suatu *item* [4]. Misal dalam suatu dokumen atribut pembentuknya yaitu *term* yang ada dalam dokumen tersebut. Dalam penelitian jurnal [5] dijelaskan bahwa metode *Content-Based Filtering* memiliki 2 teknik umum dalam membuat proses rekomendasi salah satunya *heuristic-based* yang di dalamnya menggunakan TF-IDF (*term frequency-invers document frequency*). Metode ini bersifat *user independence* [6], tidak akan bergantung pada situasi dimana *item* merupakan *item* baru (belum pernah dipilih oleh pengguna manapun) maupun bukan *item* baru sehingga ketika pengguna akan mencari *wedding organizer* pada kategori tertentu maka sistem akan mencoba merekomendasikan list *wedding organizer* dengan kategori yang juga berada di *wedding organizer* lainnya.

Penerapan algoritma *Random Forest Regression* sendiri dalam penelitian adalah untuk menyempurnakan proses rekomendasi.yang sebelumnya diproses oleh *Content-Based Filtering. Random Forest* adalah pengembangan dari Algoritma C4.5 (*decision tree*) dimana pada setiap *decision tree* dilakukan proses *training* data menggunakan sampel individu dan setiap atribut dipecah pada *tree* yang dipilih antara atribut subset yang bersifat acak [7]. Diketahui dari (Breiman, 2001) *Random Forest* juga bagian pengembangan dari metode CART, yaitu dengan menerapkan metode *bootstrap aggregating (bagging).* Setiap pohon ditanam dengan cara yang sama. Beberapa fungsi pembelajaran dari hasil *Random Forest* menggunakan strategi *ensemble bagging* untuk mengatasi *overfitting* apabila berhadapan dengan data *train* yang kecil [8]. Diketahui juga tingkat akurasi dari algoritma klasifikasi *Random Forest* cukup baik [9]. Seperti pada penelitian terkait penerapan *Random Forest Regression* milik Mega Luna diketahui memiliki nilai MAPE sebesar 23% dan nilai akurasi sebesar 76,7% untuk penerapan *Random Forest* pada sistem prediksi harga kamera bekas dimana ketika nilai MAPE semakin rendah berarti tingkat kesalahan dari hasil model tersebut semakin kecil maka nilai akurasi dari model semakin baik [10]. Penelitian lagi yang melakukan prediksi yaitu prediksi jumlah mahasiswa baru dengan menggunakan model *Random Forest* menerangkan bahwa *Random Forest* juga sangat baik dalam memecahkan masalah klasifikasi dan regresi dengan mendapat hasil akurasi 99,8% dan nilai MSE serta MAE yang sangat baik yaitu di bawah 5% [11].

Dua metode digunakan dengan tujuan memberikan hasil rekomendasi sesuai keinginan pengguna berdasarkan lokasi dan harga yang diinputkan. Jadi, kerja metode *Content Based Filtering* pada penelitian ini akan mencari kesamaan kalimat dalam fitur deskripsi fasilitas melalui cosine similarity yang terbentuk dan setelahnya itu akan dihitung oleh *Random Forest Regression*. Maka dari itu, di dalam artikel penulis akan membuat sistem rekomendasi dalam penelitian berjudul "Sistem Rekomendasi Wedding Organizer Menggunakan Metode *Content-Based Filtering* Dengan Algoritma *Random Forest Regression*". Diharapkan dalam sistem ini dapat membantu masyarakat umum lebih mudah mendapat penyedia jasa pernikahan yang sesuai dengan kriteria mereka.

II. METODOLOGI PENELITIAN

Tahapan dari metode penelitian untuk meciptakan Sistem Rekomendasi Wedding Organizer Menggunakan Metode *Content-Based Filtering* dengan Algoritma *Random Forest Regression* akan ditunjukkan pada Gbr 1.

Gbr. 1 Tahapan Penelitian

Penelitian ini memiliki beberapa tahap pada prosesnya supaya menghasilkan suatu penelitian yang berhasil seperti

diuraikan pada Gbr. 1. Adapun penjelasan dari masing-masing tahapan akan dijelaskan sebagai berikut:

A. *Identifikasi Masalah*

Tahap pertama dalam penelitian merupakan identifikasi masalah yaitu penulis menganalisis permasalahan yang biasa terjadi ketika akan melangsungkan pernikahan. Permasalahan tersebut yang akan menjadi dasar pertimbangan dalam menentukan *wedding organizar* yang sesuai dengan kesanggupan biaya dan kebutuhan mereka. Meninjau dari permasalahan yang biasa terjadi maka akan lebih baiknya ada sebuah sistem rekomendasi untuk *wedding organizar* agar tidak sampai terjadi kesalahan bagi calon pengantin dalam memilih penyedia jasa pernikahan untuk membantu mereka mempersiapkan segalanya.

B. *Penentuan Algoritma Penelitian*

Tahap kedua yaitu penentuan algoritma penelitian. Dalam penelitian membangun sistem rekomendasi disini mengangkat metode *Content-Based Filtering* dan algoritma *Random Forest Regression*.

C. *Pengumpulan Dataset*

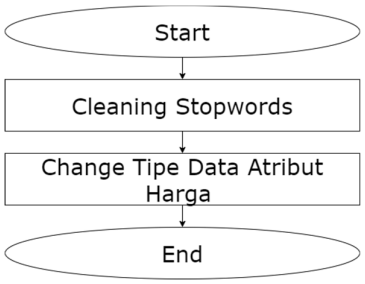
Tahapan selanjutnya adalah pengumpulan dataset. Data yang dikumpulkan diperoleh dari informasi beberapa *wedding organizer*. Adapun tahap pengumpulan data yang dilakukan ada 2 yaitu teknik observasi dengan mendatangi langsung *wedding organizer* di daerah Surabaya dan teknik pengumpulan data sekunder (Sugiyono, 2011) yang mana data berasal dari sumber yang tidak langsung memberikan data kepada pengumpul data seperti melalui orang lain maupun melalui dokumen terkait fasilitas berdasarkan kesesuaian harga setiap paket [12]. Data yang sudah diperoleh berisi nama beberapa *wedding organizer* yang ada di beberapa lokasi, nama paket beserta harga dan rincian fasilitas setiap paketnya, lokasi berupa alamat serta latitude dan longitudenya. Wilayah yang diambil pada pengumpulan data *wedding organizer* mencakup Kota Surabaya, Sidoarjo, Gresik, Lamongan, dan Mojokerto.

TABEL I
ATRIBUT DATASET

Atribut	Deskripsi	Tipe Data
Nama	Nama *wedding organizer*	String
Paket	Paket yang tersedia di setiap nama *wedding organizer*	String
Harga	Harga di setiap paket pernikahan dalam setiap *wedding organizer*	Float
Deskripsi Fasilitas	Deskripsi Fasilitas yang didapat dalam setiap paket	String
Alamat	Alamat dari setiap *wedding organizer*	String
Latitude	Latitude di setiap lokasi yang dicantumkan	Float
Longitude	Longitude di setiap lokasi yang dicantumkan	Float

D. *Preprocessing*

Tahap *preprocessing* adalah proses mengubah bentuk data yang sebelumnya belum terstruktur menjadi data terstruktur menyesuaikan kebutuhan untuk diproses lebih lanjut. *Preprocessing* dalam penelitian ini memiliki tujuan agar data yang digunakan dapat diproses menjadi angka dengan TF-IDF dan *cosine similarity* di metode selanjutnya. beberapa tahapan diantaranya proses *cleaning stopwords.* Proses *cleaning stopwords* adalah proses untuk menghilangkan kata-kata yang *missing* atau yang tidak diperlukan dalam deskripsi, sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata yang penting saja. Lalu, proses *change* tipe data harga dari string ke integer untuk disesuaikan dengan kebutuhan sistem. Proses *preprocessin*g dilakukan agar data bisa diproses oleh algoritma. Diagram alur proses *preprocessing* akan ditunjukkan pada Gbr. 2.



Gbr. 2 Diagram Alur *Preprocessing*

E. *Metode Content-Based Filtering*

Metode yang digunakan yaitu pendekatan *Content-based Filtering* sebagai metode dalam sistem rekomendasi yang menghasilkan rekomendasi berdasarkan atribut yang terkait dengan *item* lain. *Content-based Filtering* memberikan

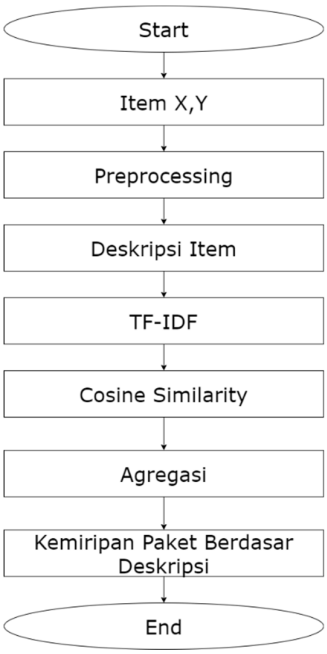
rekomendasi berdasarkan hasil analisis kemiripan item yang telah dinilai oleh penggunanya [13]. Kemiripan *item* dihitung berdasar pada fitur-fitur yang ada pada *item* yang dibandingkan. *Item* dalam penelitian diambil dari fasilitas yang tersedia dalam kolom deskripsi fasilitas. Penerapannya adalah sistem akan membuat kesamaan deskripsi dari atribut deksripsi fasilitas berdasar bobot vektornya menggunakan TF-IDF dan *cosine similarity*. TF-IDF menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam dokumen tertentu dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Proses dimulai dari menghitung *term* yang ada pada setiap dokumen (TF). Kemudian dilanjut dengan menghitung jumlah dokumen yang mempunyai *term* tertentu (DF). Setelah itu menghitung *Inverse Document Frequency* (IDF) dan yang terakhir nilai TF dikalikan dengan IDF. Rumus metode TF-IDF yang menghitung bobot setiap kata di dokumen dituliskan dengan:

$$Wdt = tf(t,d) \times IDF(t),$$

sedangkan nilai IDF didapatkan dari rumus

$$IDF(t) : log(D/df)$$

Diagram alur dari prosedur *Content-Based Filtering* akan ditunjukkan pada Gbr. 3.



Gbr. 3 Diagram Alur Metode *Content-Based Filtering*

Dari Gbr. 3 dijelaskan bahwa data yang sebelumnya sudah diproses dalam tahap *preprocessing* dilanjut untuk proses menjadi matriks angka dengan TF-IDF setelah itu kesamaannya dihitung dengan *cosine similarity*. Dalam berlangsungnya metode *Content-Based Filtering* kami juga menggunakan pendekatan agregasi untuk perbandingan sehingga menghasilkan data berupa dua paket wedding dalam satu *row* beserta *cosine similarity*. Setelah itu, terdapat hasil berupa kemiripan *string* deskripsi satu dengan yang lainnya.

F. *Metode Random Forest Regression*

Metode selanjutnya adalah Metode *Random Forest* menggunakan pendekatan *Random Forest Regression* [14]. Untuk membuat model *Random Forest* menggunakan fungsi *Random Forest Regressor* dengan parameter *bootstrap* dan n_estimators, yang mana n_estimators merupakan jumlah pohon dan *bootstrap* yang bernilai *true. Bootstrap* dan *replacement* digunakan ketika membangun pohon yang mana jika *bootstrap* memiliki nilai false maka akan menggunakan keseluruhan *dataset* untuk membangun *forest.* Performansi prediksi dari *Random Forest* sendiri diukur dari tingkat eror hasil prediksi. Untuk mengukur tingkat akurasi yaitu dengan *Mean Absolute Percentage Error* (MAPE) [15]. MAPE merupakan rata-rata dari keseluruhan persentase selisih antara data *actual* dan data hasil prediksi. Semakin rendah nilai MAPE, maka kemampuan dari model prediksi yang digunakan dapat dikatakan baik. Rumus perhitungan *Mean Absolute Percentage Error* (MAPE) adalah sebagai berikut:

$$MAPE = \frac{\sum_t^n = 1 \left| \left(\frac{A_t - F_t}{A_t} \right) 100 \right|}{\text{n}}$$

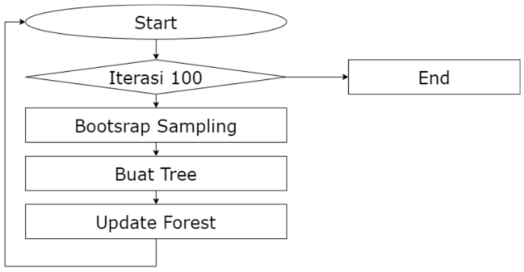
Keterangan:

At = Nilai harga asli

Ft = Hasil prediksi

N = Besar data prediksi

Proses *Random Forest* dimulai dari setiap distribusi prediksi diperoleh melalui pencarian *top-down* dari setiap pohon keputusan *random forest*. Setelah itu vektor kumulatif dihitung melalui akumulasi distribusi prediksi semua pohon. Jadi, masing-masing *tree* akan mengambil beberapa dari fitur secara *random*. Dari setiap *tree* akan dihitung hasilnya, lalu hasil total dari rata-rata setiap *tree* ditambah ke *forest* dan diiterasi sebanyak 100 kali. Setelah itu akan dihasilkan prediksi. Untuk evaluasi tiap model *Random Forest* digunakan *cross validation* yang berfungsi untuk mengevaluasi serta membandingkan beberapa algoritma dalam satu proses dengan memilah data menjadi 2 yaitu digunakan untuk melatih model dan menguji model [16]. Salah satu susunan *cross validation* adalah *K-fold cross validation* dimana pada prosesnya data dipilah menjadi beberapa *fold*. Tujuan dari validasi *K-fold* adalah untuk memvalidasi algoritma *Random Forest* agar lebih teruji dan mendapat hasil kinerja yang valid [17]. Diagram alur dari prosedur *Random Forest* akan ditunjukkan pada Gbr. 4.



Gbr. 4. Diagram Alur Metode *Random Forest*

G. *Pengembangan Sistem*

Pengembangan sistem merupakan tahap dimana sistem yang dibuat akan dirancang menuju pada tujuan dari pembuatan artikel ilmiah ini sehingga benar benar menjadikan sebuah sistem rekomendasi. Pada tahap pengembangan sistem terdapat analisis sistem dan desain sistem.

1) *Analisis Sistem*

a. Identifikasi Masalah

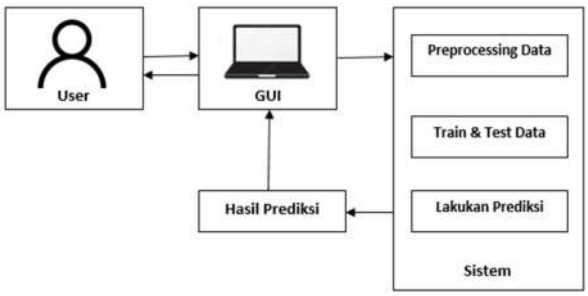
Berdasar latar belakang serta pembahsan sebelumnya, maka permasalahan yang dapat diuraikan yaitu bagaimana merealisasikan Sistem Rekomendasi *Wedding Organizer* Menggunakan Metode *Content-Based Filtering* dengan Algoritma *Random Forest Regression*

b. Analisis Kebutuhan Fungsional

Sistem Rekomendasi *Wedding Organizer* dalam penelitian ini dapat melakukan beberapa hal seperti:

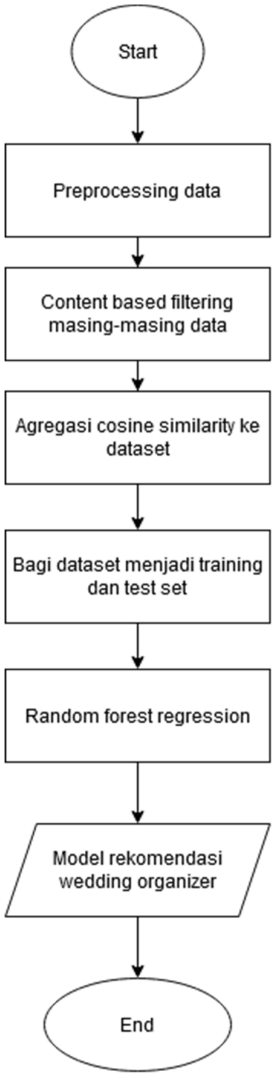
- Pengguna dapat mencari rekomendasi *wedding organizer* dengan memasukkan harga
- Pengguna dapat mencari rekomendasi *wedding organizer* dengan menyertakan lokasi melalui saran pencarian
- Pengguna dapat memilih lokasi dengan meng-klik *location* di maps
- Sistem dapat menampilkan empat hasil rekomendasi dengan persentase tertinggi

2) *Desain Sistem*



Gbr. 5. Desain Sistem

Dari Gbr. 5 dijelaskan bahwasanya proses awal dengan pengguna diarahkan ke GUI atau tampilan program yang berisi *input* harga dan lokasi. Setelah pengguna memasukkan data, sistem akan memproses data masukan pengguna dengan model *Random Forest Regression* yang telah dilatih. Model ini akan menghasilkan daftar perhitungan *cosine sismilarity* antara masukan pengguna dan 205 data yang ada dalam dataset. Lalu akan diambil 4 data dengan *similarity* tertinggi untuk ditampilkan kepada pengguna sebagai hasil rekomendasi paket *wedding organizer* berdasarkan masukan pengguna tadi. Dan untuk bagian dari berjalannya proses sistem ada di Gbr. 6. Berikut diagram alur atau *flowchart* dari proses sistem.



Gbr. 6. *Flowchart* proses sistem

Gbr. 6 merupakan diagram alur yang menggambarkan proses dari sistem yang dijalankan oleh *user*. Jadi, proses berlangsung dengan diawali *import* dataset yang akan diolah dahulu pada tahap *preprocessing* yaitu proses *cleaning stopwords* pada atribut deskripsi fasilitas dan perubahan tipe data pada atribut harga. Setelah selesai dari *preprocessing,* proses pengujian dilanjutkan dengan menghitung TF-IDF dari atribut deskripsi fasilitas lalu didapatkan *cosine similarity* dari kesamaan deskripsi fasilitasnya. Setelah itu dataset menjadi harga, lokasi, dan *cosine similarity* lalu diagregasi. Dataset yang sudah memiliki *cosine similarity* diproses dengan *Random Forest Regression* dimana harga dan lokasi menjadi fitur dan *cosine similarity* menjadi target, lalu diregresi. Data

dibagi dua menjadi data latih dan data uji. Dalam *Random Forest* kita menentukan jumlah pohon yang akan digunakan dalam *Random Forest* dan akan ada pengukuran kinerja model *random forest* atau dievaluasi dengan 10-*fold cross validation*, sehingga model yang telah terbentuk bisa langsung diuji dengan 10 kali pengujian.

III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan Sistem Rekomendasi Wedding Organizer menggunakan Metode *Content-Based Filtering* dengan Algoritma *Random Forest Regression* dimana prosesnya terbagi menjadi 4 yaitu *preprocessing,* proses dari metode *Content-Based Filtering*, proses *Random Forest* dan pengujian sistem. Berikut adalah penjelasan hasil dari 4 proses tersebut:

A. *Hasil Preprocessing*

Dataset yang diperoleh masih berupa data mentah berisi 205 data nama *wedding organizer*, paket, harga, deskripsi fasilitas, alamat, kota, latitude, longitude yang mana masih perlu perbaikan pada beberapa atributnya agar data dapat diproses oleh algoritma. Data yang masih mentah memiliki beberapa kolom yang perlu diubah, masih memiliki tipe yang belum sesuai dengan kebutuhan sistem seperti sebagai berikut:

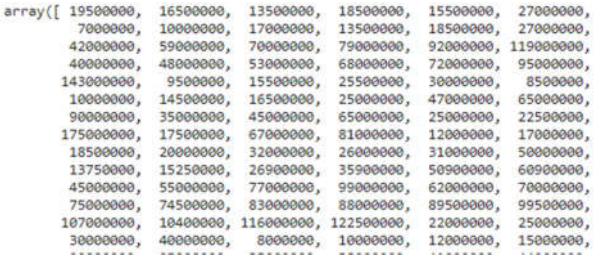
TABEL III
DATA YANG BELUM DIPROSES

No	Nama	Paket	Harga	Deskripsi_ Fasilitas	Alamat	Kota	Latitude	Longitude
1	Zulfania Wedding	Paket Gedung Gradiel	IDR 13.500.000	Dekorasi Pelaminan 6M + Taman + Lampu + Karpet + Bunga asli,Pelengkung Pintu Masuk/Pregola	Jl. Simo Gn. Bar. Tol II No.6 B, Simomulyo	Surabaya	-7.270632	112.705787

Dari Tabel II diketahui bahwa masih ada data yang perlu diperbaiki dan masih ada tipe data dari salah satu atribut yang belum sesuai sehingga perlu dilakukan pemrosesan data awal dengan google colaboratory notebook. Proses pertama dari preprocessing ialah cleaning stopwords. Berikut hasil dari preprocessing tahap cleaning stopwords ditunjukkan pada Gbr. 7.

Gbr. 7. Pre-processing tahap stopwords

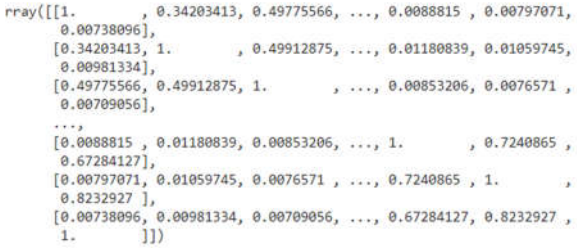
Setelah itu tahap mengganti tipe data dalam atribut harga dari string menjadi integer. Berikut hasil dari preprocessing tahap ubah tipe data ditunjukkan pada Gbr. 8.



Gbr. 8. Pre-processing tahap ubah tipe data atribut harga

B. *Hasil Content-Based Filtering*

Fungsi metode *content-based filtering* dalam sistem yaitu .menghitung TF-IDF yang menghasilkan matriks lalu dari setiap matriks tersebut dibandingkan dan dicari cosine similarity sehingga menghasilkan similaritas. Setelah menghasilkan matriks similarity dilanjut proses agregasi ke dataset yang menghasilkan 2 paket wedding organizer beserta cosine similarity dalam satu baris. Berikut ini hasil dari proses Content-Based Filtering tahap perhitungan TF-IDF pada Gbr. 8.



Gbr. 9. Hasil Content-Based Filtering tahap TF-IDF

Setelah perhitungan TF-IDF untuk menemukan *cosine similarity*, lalu akan dilakukan agregasi *cosine similarity* sesuai dengan paket wedding organizernya. Proses agregasi mengahsilkan 42025 baris data berupa pasangan satu paket *wedding organizer* dengan paket *wedding organizer* lainnya beserta *cosine similarity* dari dua paket yang dibandingkan. Data ini kemudian akan digunakan pada proses *Random Forest*. Berikut hasil agregasi dari proses *Content-Based Filtering* pada Gbr. 9.

Gbr. 10. Content-Based Filtering tahap agregasi

C. *Hasil Random Forest Regression*

Atribut yang akan digunakan data hasil agregasi adalah HARGA_1, LATITUDE_1, LONGITUDE_1, HARGA_2, LATITUDE_2, dan LONGITUDE_2. Sedangkan atribut COS_SIMILARITY akan digunakan sebagai target. Dalam proses *Random Forest Regression* data agregasi yang berjumlah 42025 akan dibagi menjadi 2 bagian, yaitu data *training* sebesar 80% dengan jumlah data 33620 dan data *testing* sebesar 20% dengan perbandingan jumlah data uji yang lebih kecil dibanding data latih yaitu berjumlah 8405 data.

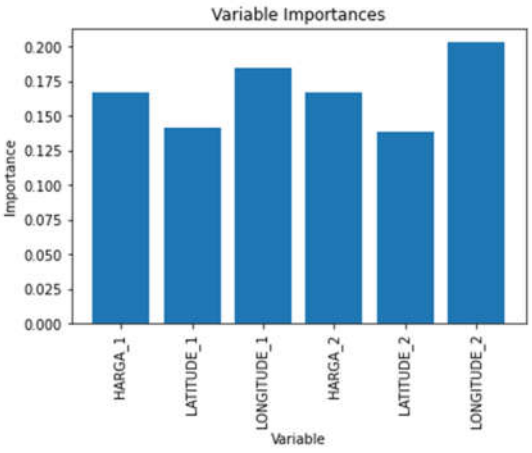
Setelah itu pengujian *cross validation* menentukan jumlah K dari beberapa variasi K yang digunakan hingga mendapatkan akurasi terbaik. Tahap ini perlu dilakukan untuk mendapatkan model *Random Forest* disini yang menggunakan 100 *tree* dengan pengujian K yang mempunyai nilai error yang paling rendah, yang berarti bahwa model tersebut mempunyai nilai akurasi yang paling tinggi. Pada pengujian K ditentukan nilai variasi K dari 2-fold hingga 10-*fold*. Nilai error dihitung menggunakan Mean Absolute Percentage Error (MAPE). Penggunaan MAPE dikarenakan untuk menunjukkan nilai error dalam persentase sehingga mudah dipahami dan dibandingkan. MAPE juga sering digunakan dalam evaluasi *forecasting*. Berikut ini merupakan Tabel hasil pengujian dari masing-masing K-*fold* cross validation yang digunakan pada pengujian metode *Random Forest Regression.*

TABEL III
HASIL PENGUJIAN K-FOLD DARI 10-FOLD

K	MAPE (%)	Akurasi (%)
2	27.925 %	72.074 %
3	21.784 %	78.215 %
4	19.345 %	80.654 %
5	18.231 %	81.768 %
6	17.475 %	82.524 %
7	17.027 %	82.972 %
8	16.631 %	83.368 %
9	16.320 %	83.679 %
10	16.249 %	83.750 %

Berdasarkan pengujian dengan menggunakan banyak jumlah *tree*, model *Random Forest Regression* dengan nilai K=10 menghasilkan nilai akurasi paling tinggi (83,750 %). Sehingga model yang akan digunakan adalah *Random Forest* dengan K=10. Model ini yang akan diberikan proses *training* dan disimpan untuk digunakan dalam perhitungan rekomendasi di aplikasi web.

Tahap selanjutnya adalah mengukur kepentingan variabel yang digunakan terhadap nilai *cos similarity* berdasarkan nilai *feature importance*.



Gbr. 11. Diagram *feature importance*

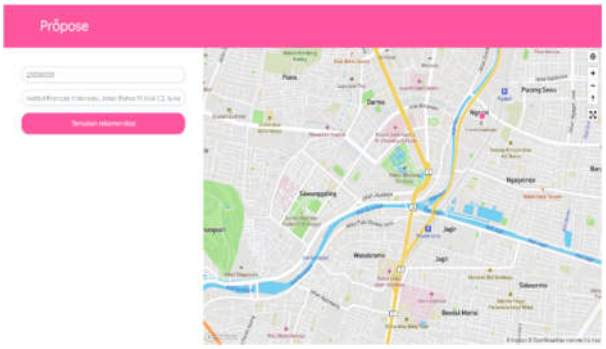
Dari diagram *feature importance* dapat dilihat bahwa variabel dengan nilai importance tertinggi adalah LONGITUDE_2 (0.200). Meskipun begitu dalam diagram dapat dilihat bahwa semua variabel mempunyai nilai *importance* yang hampir sama sehingga semua variabel bisa dikatakan cukup berpengaruh terhadap hasil perhitungan *similarity* antara dua paket *wedding organizer*.

D. *Hasil Pengujian Sistem*

Sistem dibentuk dengan menggunakan Bahasa pemrograman python. Setelah model terbentuk maka dibuat tampilan antarmuka yang dapat menghubungkan pengguna (*user*) dengan proses yang berlangsung. *Software Pycharm* menjadi platform yang penulis gunakan untuk mengembangkan aplikasi rekomendasi *wedding organizer* yang berbasis python. Berikut adalah tampilan dari halaman beranda sistem rekomendasi *wedding organizer*:

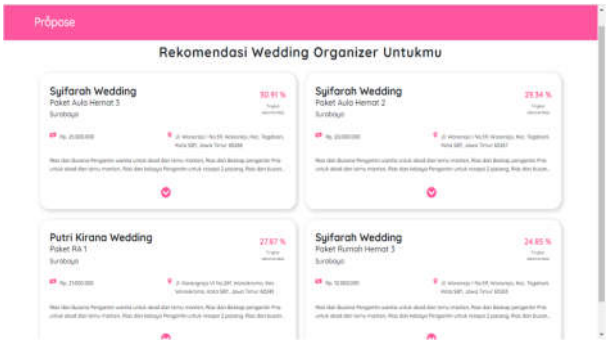
7Gbr. 12. Beranda Sistem Wedding Organizer

Pada tampilan sistem di Gbr. 12 kita mulai memasuki beranda sistem rekomendasi dimana terdapat *button* Mulai Temukan yang akan membawa *user* menuju halaman selanjutnya yaitu halaman proses *input* pada Gbr. 13.



Gbr. 13. Proses Input Sistem Wedding Organizer

Pada tampilan sistem di Gbr. 13 merupakan halaman proses rekomendasi berlangsung dimana *user* dapat memasukkan nominal harga serta lokasi yang diinginkan. Setelah itu *user* akan mendapatkan hasil rekomendasi seperti yang terdapat pada Gbr. 14. Berikut ini adalah tampilan dari halaman hasil rekomendasi *wedding organizer* sesuai dengan inputan pengguna.



Gbr. 14. Tampilan Hasil Rekomendasi Wedding Organizer

Pada tampilan hasil rekomendasi di Gbr. 14 ditunjukkan hasil rekomendasi *wedding organizer* berdasarkan inputan pengguna. Beberapa hasil pengujian rekomendasi juga ditunjukkan pada tabel IV berikut.

TABEL IV
HASIL PENGUJIAN SISTEM REKOMENDASI

No	Input		Hasil		
	Harga	Location	Wedding	Harga	Lokasi
1.	Rp. 50.000.000	Waru	Juwita Wedding	Rp. 52.000.000	Jl. Rungkut Mejoyo Selatan V No.19
			Juwita Wedding	Rp. 37.000.000	Jl. Rungkut Mejoyo Selatan V No.19
			Juwita Wedding	Rp. 80.000.000	Jl. Rungkut Mejoyo Selatan V No.19
			Qaia Wedding	Rp. 82.900.000	Jl. Samanhudi No. 64, Kroman, Pakelingan, Gresik
2.	Rp. 40.000.000	Alun-Alun Gresik, Jl. Pahlawan, Gresik	Kusuma Ayu Wedding	Rp. 44.500.000	Bp Wetan,Jl. Seruni No. 7, Sumberejo, Gresik
			Jaehan Wedding	Rp. 44.900.000	Jl. Sumatra No. 37, Setinggi, Randuagung, Gresik
			Jaehan Wedding	Rp. 49.900.000	Jl. Sumatra No. 37, Setinggi, Randuagung, Gresik
			Diva Wedding	Rp. 41.300.000	Jl. Gembili Raya No. 5, Jagir, Wonokromo, Surabaya
3.	Rp. 30.000.000	BG Junc-tion, Bubu-tan, Surabaya	NDD Wedding	Rp. 27.200.000	Jl. Simo Rejo II No. 66, Simomulyo, Surabaya
			NDD Wedding	Rp. 19.000.000	Jl. Simo Rejo II No. 66, Simomulyo, Surabaya
			NDD Wedding	Rp. 34.200.000	Jl. Simo Rejo II No. 66, Simomulyo, Surabaya
			NDD Wedding	Rp. 35.200.000	Jl. Simo Rejo II No. 66, Simomulyo, Surabaya
4.	Rp. 15.000.000	Banjar-sari, Cerme, Gresik	Kusuma Ayu Wedding	Rp. 13.000.000	Bp Wetan,Jl. Seruni No. 7, Sumberejo, Gresik
			Jaehan Wedding	Rp. 26.500.000	Jl. Sumatra No. 37, Setinggi, Randuagung, Gresik
			Syifarah Wedding	Rp. 15.000.000	Jl. Wonorejo I No. 59, Wonorejo, Surabaya
			Ben Bagoes Wedding	Rp. 13.900.000	Jl. Asem Mulya No. 81-83. Asem Rowo, Surabaya
5.	Rp. 25.000.000	Sidoto-po, Suraba-ya	Citra Keraton Wedding	Rp. 23.000.000	Jl. Sidotopo Jaya 7 No. 21, Sidotopo, Surabaya
			Citra Keraton Wedding	Rp. 16.700.000	Jl. Sidotopo Jaya 7 No. 21, Sidotopo, Surabaya
			Citra Keraton Wedding	Rp. 11.000.000	Jl. Sidotopo Jaya 7 No. 21, Sidotopo, Surabaya
			Zulfania Wedding	Rp. 18.500.000	Jl. Simo Gn. Bar. Tol II No. 6B, Simomulyo, Surabaya

Dari tampilan table IV terdapat 4 hasil rekomendasi *wedding organizer* berdasarkan cosine similarity tertinggi. Informasi dari hasil rekomendasi berisi nama *wedding organizer,* nama paket, lokasi *wedding organizer,* harga paket, dan deskripsi fasilitas dari paket. Terdapat persentase tingkat rekomendasi dari masing masing *wedding organizernya* yang merupakan hasil prediksi cosine similarity terhadap inputan pengguna.

IV. KESIMPULAN

Dari penelitian yang dilakukan dan berdasarkan pembahasan sebelumnya diperoleh kesimpulan sebagai berikut:

1. Pembuatan Sistem Rekomendasi *Wedding Organizer* telah dilakukan dengan menggunakan Metode *Content-Based Filtering* dan Algoritma *Random Forest Regression*.
2. Dari penelitian menggunakan evaluasi *cross validation* 10-*fold* yang digunakan untuk *random forest* menunjukkan bahwa *random forest* dengan *tree* 100 mempunyai hasil yang terbaik dengan akurasi 83,750% dan MAPE sebesar 16,249%.
3. Sistem Rekomendasi *Wedding Organizer* Menggunakan Metode *Content-Based Filtering* dan Algoritma *Random Forest Regression* dapat membantu memudahkan para pengguna mendapatkan penyedia jasa pernikahan sesuai dengan *budget* yang dimiliki dan lokasi yang diinginkan.

V. SARAN

Berdasarkan penelitian yang dilakukan tidak dipungkiri bahwa pasti ada kekurangan, sehingga perlu adanya saran. Adapun saran yang diberikan oleh penulis sebagai berikut:

1. Sistem Rekomendasi bisa dikembangkan lagi menggunakan metode lain dengan tujuan mungkin akan mendapat hasil akurasi yang lebih tinggi atau akurat.
2. Memperbanyak pengumpulan data dalam dataset karena jika variasi data lebih banyak maka akan memperkuat hasil akurasi.

UCAPAN TERIMA KASIH

Terimakasih penulis ucapkan yang pertama kepada Tuhan Yang Maha Esa, karena atas karunia-Nya penulis dapat menyelesaikan penelitian ini dengan baik. Kemudian terimakasih ini juga ditujukan kepada kedua orang tua, dan semua pihak yang terlibat dalam terselesaikannya penelitian ini.

REFERENSI

[1] Malvyandie, H. (2017, Januari 23). *Bridestory Rilis Laporan Tren Pernikahan 2017 di Indonesia*. Retrieved from Tribunnews: https://www.tribunnews.com/tribunners/2017/01/23/bridestory-rilis-laporan-tren-pernikahan-2017-di-indonesia

[2] Utomo, B. T. W., & Anggriawan, A. W. (2015). Sistem Rekomendasi Paket Wisata Se-Malang Raya Menggunakan Metode Hybrid Content Based Dan Collaborative. Jurnal Ilmiah Teknologi Informasi Asia, 9(1), 6–13

[3] Februariyanti, H. D. (2021). IMPLEMENTASI METODE COLLABORATIVE FILTERING UNTUK SISTEM REKOMENDASI PENJUALAN PADA TOKO MEBEL Diterima : Diterbitkan : IX(I), 43–50.

[4] Badriyah, T., Fernando, R., & Syarif, I. (2018). Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. Konferensi Nasional Sistem Informasi, 1(1), 554–559.

[5] Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. Teknika, 8(1), 1–10. https://doi.org/10.34148/teknika.v8i1.139

[6] Mondi, R. H., Wijayanto, A., & Winarno. (2019). Recommendation System With Content-Based Filtering Method For Culinary Tourism In Mangan Application. *ITSMART*, 2301-7201.

[7] Frastian, N. (2018). Implementasi Komparasi Algoritma Klasifikasi Menentukan Kelulusan Mata Kuliah Algoritma Universitas Budi Luhur. Jurnal String, 2549-2837.

[8] Breiman, L. (2001). Random Forest. Machine Learning, 5-32.

[9] Penjualan, R., & Aliexpress, G. (2018). Komparasi algoritma c4.5 dengan random forest untuk rekomendasi penjualan gaun aliexpress.com. 49.

[10] Suliztia, M. L. (2020). Penerapan Analisis Random Forest pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask. Fakultas Matematika Dan Ilmu Pengetahuan Alam, 1–107.

[11] Rianto, M., & Yunis, R. (2021). Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest. Paradigma - Jurnal Komputer Dan Informatika, 23(1). https://doi.org/10.31294/p.v23i1.9781

[12] Sugiyono. (2011). Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung: Alfabeta.

[13] Puspaningtyas, S. A. (2010, Mei 22). Ssitem Rekomendasi Nilai Mata Kuliah Menggunakan Metode Content-Based Filtering. Seminar Nasional Informatika, pp. A90-A94

[14] Samudra, A. Y. (2019). Pendekatan Random Forest Untuk Model Peramalan Harga Tembakau Rajangan Di Kabupaten Temanggung. Karya Ilmiah, 1-98.

[15] Saadah, S., & Salsabila, H. (2021). Jurnal Politeknik Caltex Riau Prediksi Harga Bitcoin Menggunakan Metode Random Forest (Studi Kasus: Data Acak Pada Awal Masa Pandemic Covid-19). Jurnal Komputer Terapan, 7(1), 24–32. https://jurnal.pcr.ac.id/index.php/jkt/

[16] Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross Validation. In L. Liu, & M. T. Özsu, Encyclopedia of Database Systems (pp. 532-538). Arizona: Springer US.

[17] HULU, S. S. U. (2020). Analisis Kinerja Metode Cross Validation Dan K-Nearest Neighbor Dalam Klasifikasi Data. Universitas Sumatera Utara, 4–16.