Universitas Ahmad Dahlan

Jurnal Ilmiah Matematika
Vol. 11, No. 2, Oktober 2024, pp. 79-86
ISSN 2774-3241
http://journal.uad.ac.id/index.php/Konvergensi

Analisa dan Prediksi Cost Pada Food Mart Menggunakan Model Algoritma Random Forest Regression

Zulfatin Nafisah[a.1], Aris Thobirin[a.2]

[a] Prodi Matematika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan, Yogyakarta
[1] zulfatin2015015025@webmail.uad.ac.id
[2] aris.thobi@math.uad.ac.id

Received: *Revised:* *Accepted:*

KATAKUNCI

CFM, *Cost Acquisition Customer* (CAC), Data Mining, Random Forest Regression, Machine learning, Artificial Intelligence

ABSTRAK

Perusahaan *Convenient Food Mart* (CFM) berada di Amerika Serikat yang menjual berbagai produk bahan makanan, minuman ringan hingga makanan siap saji menerapkan strategi *Cost Acquisition Customer* (CAC) untuk mengetahui analisa target dan besaran biaya yang akan dikeluarkan sehingga tidak mengeluarkan biaya anggaran yang tinggi dan tetap mempertahankan pelanggan serta menarik pelanggan yang baru. Oleh karena itu, penulis memprediksi biaya akuisisi pelanggan tersebut menggunakan model Random Forest Regression. Berdasarkan model algoritma tersebut diperoleh nilai akurasi atau ^{2} score sebesar 0.901893 sehingga model algoritma tersebut memiliki performa model atau nilai keakuratan yang cukup baik. Sedangkan untuk feature importance atau variabel terpenting dari model algoritma tersebut terdiri dari *promotion name* dengan nilai 0.5, *store city* dengan nilai 0.2, dan *store state* dengan nilai 0.19.. Pada algoritma Random Forest Regression juga diperoleh nilai prediksi yang tidak berbeda jauh dengan nilai aktualnya sehingga besaran biaya yang dikeluarkan tidak berbeda jauh dari aslinya untuk mencapai target tertentu.

KEYWORDS

CFM, *Cost Acquisition Customer* (CAC), Data Mining, Random Forest Regression, Machine learning, Artificial Intelligence

ABSTRACT

A Convenient Food Mart (CFM) company located in the United States that sells various food products, soft drinks to ready-to-eat food implements the Cost Acquisition Customer (CAC) strategy to find out target analysis and the amount of costs to be incurred so as not to incur high budget costs and retain customers and attract new ones. Therefore, the authors predict the customer acquisition costs using the Random Forest Regression model. Based on the algorithm model, an accuracyvalue or ^{2} score of 0.901893 is obtained so that the algorithm model has a fairly good model performance or accuracy value. As for feature importan or the most important variable of the algorithm model consists of promotion name with a value of 0.5, store city with a value of 0.2, and store state with a value of 0.19. In the Random Forest Regression algorithm, the predicted value is also obtained which is not much different from the actual value so that the amount of costs incurred is not much different from the original to achieve a certain target.

Pendahuluan

Bisnis merupakan salah satu kegiatan atau aktivitas yang memberikan keuntungan pada individu maupun kelompok. Persaingan bisnis menjadi semakin ketat karena banyaknya pesaing yang muncul sehingga setiap pelaku usaha bisnis dituntut untuk memiliki inovasi dengan mengerahkan seluruh potensi yang ada untuk dapat bertahan dalam persaingan bisnis. Dalam hal persaingan bisnis penting bagi perusahaan untuk merancang strategi agar mendapatkan pelanggan sebanyak-banyaknya. *Convenient Food Mart* (CFM) adalah jaringan toko serba yang ada di Amerika Serikat, yang berdiri pada tahun 1958 di Chicago, Illinois. Kantor pusat perusahaan swasta ini berada di Mentor, Ohio, dan untuk saat ini terdapat sekitar 325 toko berlokasi di AS. CFM ini beroperasi pada sistem waralaba atau kerja sama dalam bidang usaha dengan bagi hasil sesuai dengan kesepakatan perusahaan. Pada tahun 1988 CFM menjadi jaringan terbesar ketiga di wilayah tersebut. Carden & Cherry mengiklankan CFM dengan karakter Ernest pada 1980-an. CFM menjual berbagai produk bahan makanan, minuman ringan, hingga makanan siap saji. [1]

Customer acquisition adalah proses untuk mendatangkan *customer* atau pelanggan baru ke bisnis suatu organisasi. Tujuannya untuk menciptakan strategi *acquisition* yang sistematis dan berkelanjutan sehingga dapat beradaptasi mengikuti tren perubahan. *Customer acquisition* merupakan strategi yang direkomendasikan untuk diterapkan saat memulai usaha bisnis agar memperkenalkan produk usaha kepada publik dan membuat calon pelanggan tertarik untuk menggunakan produk tersebut. *Customer acquisition* bisa dikatakan memakan biaya yang mahal sehingga apabila diterapkan tanpa strategi yang tepat, maka akan menimbulkan kerugian [2]. Berdasarkan masalah tersebut, maka dibutuhkan teknologi dan informasi untuk mengolah data. Data tersebut akan kita olah menjadi pengetahuan sebagai acuan dalam membaca dan mengetahui pola pendekatan tersembunyi dari kumpulan data, melakukan analisis tentang pengelompokkan antara data dan atribut untuk mendukung pengambilan keputusan serta pembuatan kebijakan dalam memberikan informasi mengenai analisa target dan besaran biaya yang akan dikeluarkan dalam menerapkan strategi ini yang biasa disebut dengan *Customer Acquisition Cost* (CAC).

Data *Customer Acquisition Cost* (CAC) diolah menjadi sebuah pengetahuan agar dapat bermanfaat bagi perusahaan dengan mengubah menjadi pengetahuan sehingga dapat dilakukan suatu prediksi dan estimasi tentang apa yang akan terjadi ke depan. Oleh karena itu, perlu adanya proses yang menggunakan Teknik statistik, matematik, kecerdasan buatan *(Artificial Intelligence)* dan *Machine Learning* untuk mengekstrak pengetahuan atau menentukan pola dari suatu data yang besar. Data mining suatu proses mencari pola atau informasi dalam kumpulan data yang terpilih dengan menggunakan Teknik atau metode tertentu. Data mining terbagi menjadi 5 bagian menurut peran utamanya yaitu estimasi, prediksi klasifikasi, clustering, dan asosiasi. Teknik pengolahan data mining yang sering digunakan yaitu klasifikasi. Klasifikasi yaitu proses memetakan data kedalam kelompok atau kelas yang telah ditentukan. Pendekatan data mining yang dapat diterapkan dalam melakukan tahapan penelitian ini yaitu CRISP-DM. Banyak model prosedural dan upaya untuk menstandarisasi proses penambangan data yang telah dilakukan salah satunya yaitu CRISP-DM.[10]

Tujuan penelitian ini adalah untuk memprediksi biaya akuisisi pelanggan atau yang dikenal dengan CAC *(Customer Acquisition Cost)* dengan membuat model prediksi menggunakan algoritma random forest regression. Hasil prediksi tersebut berupa output yaitu biaya akuisisi customer yang harus dikeluarkan oleh perusahaan berdasarkan nilai inputnya sehingga penerapan *customer acquisition* tidak terkesan hanya mengeluarkan biaya tanpa mendapatkan keuntungan yang jelas dan pasti.

Metode

Pengumpulan data dilakukan untuk memperoleh data atau dokumen yang dibutuhkan dalam penelitian. Adapun metode yang dilakukan untuk mengumpulkan data dalam penelitian ini yaitu melalui:

a. Dokumentasi
Metode dokumentasi diperoleh dengan mendownload dataset pada Kaggle. Data yang digunakan yaitu prediksi biaya untuk mendapatkan pelanggan *(Cost Prediction on Acquiring Customers)* yang diperoleh dari Kaggle : https://www.kaggle.com/datasets/ramjasmaurya/medias-cost-prediction-in foodmart. Dataset tersebut terdiri dari 60k pelanggan yang memiliki pendapatan *(Income),* detail promosi *(Promotion details)*, data toko *(store data)*, data penjualan *(sales data)*, biaya media *(media cost).*

b. CRISP-DM *(Cross-Industry Standard Process for Data Mining)*
Metode yang digunakan untuk implementasi data mining yaitu menggunakan kerangka kerja CRISP-DM *(Cross-Industry Standard Process for Data Mining)* yang merupakan standar metode implementasi data mining untuk industri [2], Adapun tahapannya yaitu :

1. *Business Understanding* (Pemahaman Bisnis)
   Beberapa hal yang dilakukan pada tahap ini yaitu memahami kebutuhan serta tujuan dari sudut pandang bisnis, kemudian mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan data minig.

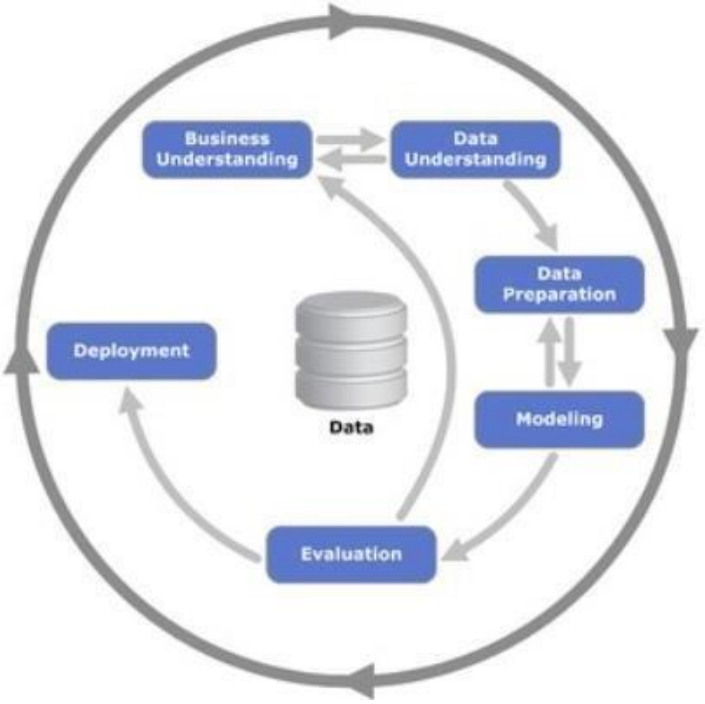
2. *Data Understanding* (Pemahaman Data)
   Tahapan ini diawali dengan mengumpulkan data, mendeskripsikan data, dan mengevaluasi kualitas data.

3. *Data Preparation* (Persiapan Data)
   Dalam tahapan ini yaitu membangun dataset akhir dari data mentah. Ada beberapa hal yang akan dilakukan diantaranya pembersihan data *(Data Cleaning)*, pemilihan data *(Data Selection),* record dan atribut-atribut serta melakukan transformasi terhadap data *(Data Transformation)* untuk dijadikan masukan dalam tahap pemodelan.

4. *Modelling* (Pemodelan)
   Pada tahapan ini secara langsung melibatkan *machine learning* untuk penentuan teknik *data mining*, alat bantu *data mining* serta algoritma *data mining*. Penelitian ini menggunakan model klasifikasi *random forest regression.*

5. *Evaluation* (Pengujian)
   Tahap ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma. Parameter yang digunakan untuk evaluasi komparansi algoritma adalah *Confusion Matrix* dengan nilai akurasi dan presisi.

6. *Deployment* (Penyebaran)
   Tahapan ini dilakukan dengan pembuatan laporan dan artikel jurnal menggunakan model yang dihasilkan.

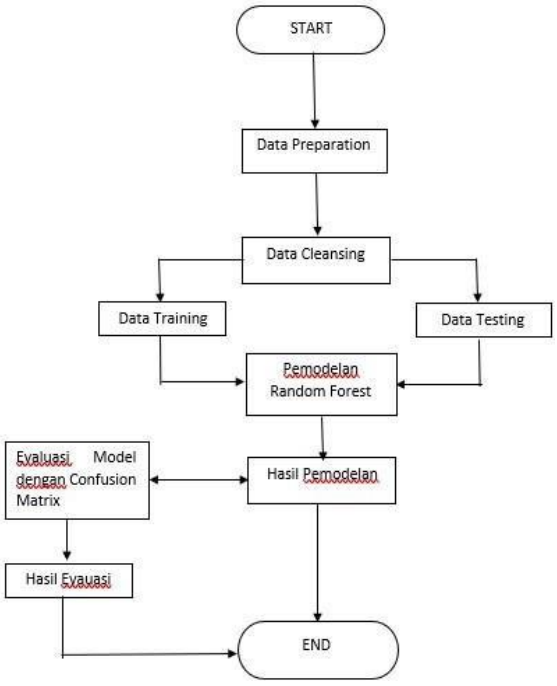


Gambar 1. CRISP-DM Process

c. Flowchart Pemodelan *Data Mining*

Flowchart atau diagram alur digunakan untuk menggambarkan alur proses mulai dari *data preparation* hingga hasil pemodelan. Diagram alur untuk menggambarkan proses pemodelan menggunakan *random forest* seperti di bawah ini :

 ISSN 2774-3241



Gambar 2. Proses Pemodelan Random Forest

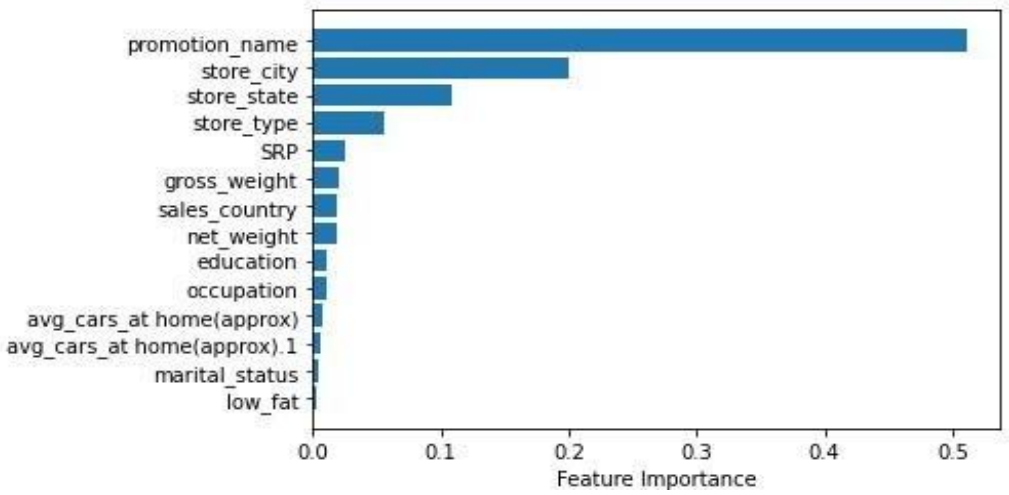
Hasil dan Pembahasan

Data yang digunakan untuk pemodelan bersumber dari dataset *media prediction and its cost.* Data tersebut berisi *customer* yang melakukan pembelian di setiap produk dari CFM. Total data awal yaitu berjumlah 60249 data dan 40 atribut. Data yang digunakan tidak mencakup semua kolom yang ada pada dataset. Beberapa variabel yang tidak digunakan akan di *dropping* atau *cleansing* data sehingga data yang digunakan menjadi 15 atribut yaitu *Promotion name, sales country, marital status, education, occupation, avg cars at home (approx), avg cars at home (approx).1, SRP, gross weight, net weight, low fat, store type, store city, store state,* dan *cost.*

Dalam modelling kita menentukan *feature importance* terlebih dahulu untuk menentukan variabel mana yang memiliki nilai terpenting. *Feature importance* model *random forest* dapat dilihat pada Tabel 1 dan untuk visualisasinya dapat dilihat pada Gambar 3.

Tabel 1. *Feature Importance*

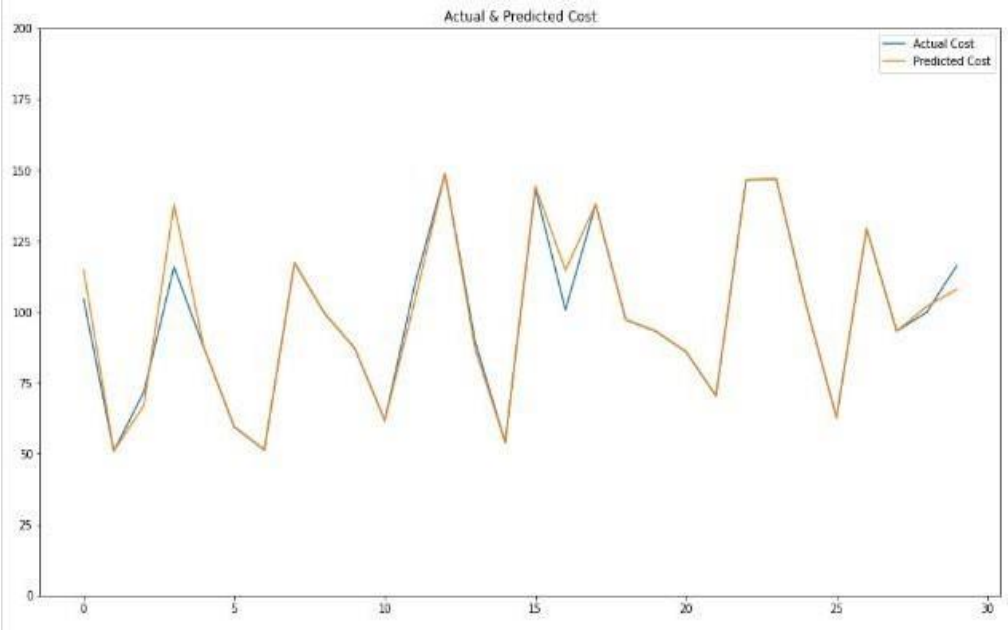
Feature	Score	Feature	Score
0	0.51143	7	0.02577
1	0.01886	8	0.01964
2	0.00474	9	0.01874
3	0.01053	10	0.00283
4	0.01052	11	0.05520
5	0.00676	12	0.19938
6	0.00649	13	0.10912



Gambar 3. *Feature importance* model random forest

Berdasarkan Tabel 1 dan Gambar 3 dari variabel yang telah diinputkan diperoleh variabel *promotion name* merupakan variabel terpenting yang memiliki skor 0.5 dan variabel *store city* dengan skor 0.2 serta variabel *store state* dengan skore 0.19.

Algoritma yang digunakan untuk prediksi pada kasus *Customer Acquisition Cost* (CAC) yaitu *random forest regression*. Berdasarkan evaluasi modelnya diperoleh nilai R^2 yaitu 0.901893 dan nilai MSE (Mean Squared Error) yaitu 88.59 sedangkan MAE (Mean Absolute Error) yaitu 3.234 sehingga *random forest regression* dapat dikatakan memiliki nilai performa yang baik dalam mengatasi prediksi dari dataset yang digunakan. Pada hasil prediksi diperoleh nilai aktual dan nilai prediksi yang dapat dilihat pada Gambar 4.



Gambar 4. Perbandingan Data Prediksi dan Data Aktu

Berdasarkan Gambar 4 pada hasil prediksi diperoleh nilai aktual dan nilai prediksi yang mempunyai akurasi atau presisi yang kuat satu sama lain. Hal ini membuktikan bahwa model algoritma *random forest* cocok pada dataset yang digunakan.

Dengan demikian, suatu data dapat dieksplorasi untuk mengetahui pengaruh variabel independent terhadap variabel dependent menggunakan *Random Forest* dengan algoritma yang tepat. Ketepatan algoritma ini menghasilkan model yang baik dengan akurasi yang tinggi sehingga nilai prediksi semakin mendekati nilai aktual. Diperoleh nilai prediksi dan nilai aktual seperti pada Tabel 2.

Tabel 2. Perbedaan Nilai Prediksi dan Nilai aktual

	Nilai Aktual	Nilai Prediksi	Perbedaan (Aktual - Prediksi)
37874	114.60	104.2770	1.032300e+01
17790	50.79	50.7900	2.842171e-14
37964	67.01	71.7161	-4.706100e+00
21346	137.77	115.7631	2.200690e+01
47214	87.07	87.0700	2.131628e-13
25348	59.40	59.3812	1.880000e-02
26950	51.27	51.2700	-7.815970e-14
39923	117.29	117.0793	2.107000e-01
16384	99.38	99.3800	1.563194e-13
18830	87.07	87.0700	2.273737e-13
20186	61.65	61.6500	9.237056e-14
56979	103.90	109.7030	-5.803000e+00
57715	148.62	148.6200	-3.126388e-13
39349	86.79	89.4245	-2.634500e+00
37241	53.82	53.8200	2.131628e-14
17977	144.18	143.2028	9.772000e-01
36393	114.60	100.6392	1.396080e+01
4464	137.70	137.7000	-3.126388e-13
8241	97.13	97.1300	7.105427e-14
36376	93.07	93.0700	2.415845e-13

Pada Tabel 2 diperoleh nilai prediksi berdasarkan data testing dan data training menggunakan perbandingan 80 % dan 20 % yang kemudian akan dibandingkan dengan nilai aktual besaran biaya yang dikeluarkan agar dapat mencapai target sesuai yang telah ditentukan ternyata tidak berbeda jauh dari nilai aslinya.

Simpulan

Dalam upaya bertahan di persaingan bisnis dan mendapatkan pelanggan baru *Convenient Food Mart* (CFM) menerapkan strategi *customer acquisition* dengan menggunakan algoritma *random forest regression* untuk memprediksi biaya yang harus dikeluarkan agar mendapatkan pelanggan yang diperoleh dari variabel-variabel yang telah diambil yaitu variabel *promotion name* yang memiliki *feature importance* dengan nilai 0.5 dan variabel *store city* dengan nilai 0.2. Dari kedua variabel tersebut *promotion name* dapat dilakukan dengan berbagai media dan berbagai cara sehingga *cos*t yang dikeluarkan juga besar. Begitu juga dengan *store city* yang juga berpengaruh terhadap *cost,* dimana *store* yang berada pada kota yang lebih maju dan modern mengakibatkan *cost* yang tinggi juga. Selain itu, terdapat juga variabel *store state* dengan nilai 0.19. Variabel ini juga memiliki hubungan dengan *cost* yang tinggi apabila *store* tersebut di negara yang maju sehingga membutuhkan biaya yang juga mahal.

Daftar Pustaka

[1] Convenient Food Mart. Diakses pada 12 Januari 2023, https://www.wikiwand.com/en/Convenient_Food_Mart

[2] Hasanah, Msy Aulia, Sopian Soim, and Ade Silvia Handayani. "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir." *Journal of Applied Informatics and Computing* 5.2 (2021): 103-108.

[3] Kurniawan, Andreas Beny. "PENDEKATAN RANDOM FOREST UNTUK MEMPREDIKSI NASABAH YANG BERPOTENSI MEMBUKA TABUNGAN DEPOSITO."

[4] Larose, Daniel T. "Data Mining: Methods and." (2006).

[5] Permana, Sendi, Rosadi Rosadi, and Nikki Nikki. "APPLICATION OF CLASSIFICATION ALGORITHM FOR SALES PREDICTION." *TEKNOKOM* 5.2 (2022): 119-124.

[6] Sangaralingam, Kajanan, et al. "High Value Customer Acquisition & Retention Modelling–A Scalable Data Mashup Approach." *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019.

[7] Suliztia, Mega Luna. "Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask." (2020).

[8] Ventura, "Costumer Acquisition Cost (CAC)," 2019. https://medium.com/@venturaofficialmedia/costumer-acquisition-cost-cac f49f69c1c324

[9] Wahyudi, Wahyudi, and Sutoyo Sutoyo. "PENERAPAN CUSTOMER ACQUISITION DALAM PERTUMBUHAN BISNIS PADA DAPOER SUPER SAMBAL PADANG." *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi* 8.2: 109-115.

[10] Yulianto, Agus. "Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes." *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer* 6.1 (2021): 41-47.