

**PREDIKSI *COST PER ACTION* (CPA) PADA IKLAN FACEBOOK
PT XYZ MENGGUNAKAN ALGORITMA RANDOM FOREST**

TUGAS AKHIR



MILLAH FATIMAH

1132001019

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER**

UNIVERSITAS BAKRIE

JAKARTA

2017

**PREDIKSI *COST PER ACTION* (CPA) PADA IKLAN FACEBOOK
PT XYZ MENGGUNAKAN ALGORITMA RANDOM FOREST**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
sarjana Komputer**



MILLAH FATIMAH

1132001019

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA**

2017


Universitas Bakrie

HALAMAN PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan benar.

Nama : Millah Fatimah

NIM : 1132001019

Tanda Tangan : 

Tanggal : 14 September 2017

Universitas Bakrie

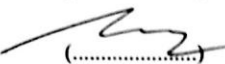
HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Millah Fatimah
NIM : 1132001019
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul Skripsi : Prediksi *Cost Per Action* (CPA) Pada Iklan Facebook PT XYZ Menggunakan Algoritma Random Forest

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

DEWAN PENGUJI

Pembimbing : Guson P. Kuntarto, S.T., M.Sc. 
(.....)
Penguji : Prof. Dr. Hoga Saragih, S.T., M.T. 
(.....)
Penguji : Dr. Siti Rohajawati, S.Kom, M.Kom. 
(.....)
Ditetapkan di : Jakarta
Tanggal : 14 September 2017

UNGKAPAN TERIMA KASIH

Alhamdulillahirrabbi'l'aalamiin, Puji dan syukur kehadiran Allah SWT karena atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir dengan judul “Prediksi *Cost Per Action* (CPA) Pada Iklan Facebook PT XYZ Menggunakan Algoritma Random Forest” ini ditulis untuk memenuhi salah satu syarat dalam menyelesaikan perkuliahan pendidikan strata satu (S1) pada Program Studi Informatika, Universitas Bakrie.

Banyak pihak yang telah membantu penulis dalam penelitian dan penulisan Tugas Akhir ini, baik itu berupa bimbingan, saran, maupun dukungan secara moril dan materil. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Bapak Hoga Saragih, S.T., M.T., selaku Kepala Program Studi Informatika, yang senantiasa memberikan masukan dan motivasi kepada penulis;
2. Bapak Guson P. Kuntarto, S.T., M.Sc., selaku dosen pembimbing, yang telah meluangkan waktunya serta memberikan bimbingan, saran, dan perbaikan dalam menyelesaikan penelitian ini;
3. Bapak Berkah I Santoso, S.T., M.T.I., selaku dosen pembahas yang memberikan saran dan perbaikan terhadap penelitian ini;
4. Ibu Dr. Siti Rohajawati, S.Kom., M.Kom., selaku dosen penguji yang memberikan saran dan perbaikan dalam penelitian ini;
5. Seluruh Bapak/Ibu dosen Program Studi Informatika UB, yang telah memberikan banyak ilmu, pengetahuan, wawasan kepada penulis selama perkuliahan;
6. Keluarga tercinta, kedua Orang tua penulis (Bapak M. Zaenudin dan Ibu Nur Hasanah) dan saudara kandung penulis (Mazlin Chairany dan Mahdy Zia Uzzaman) yang telah memberikan dukungan dan doa yang sangat berarti bagi penulis;

Universitas Bakrie

7. Mentor selama magang (Mas Yodi, Mas Wawan, Mas Rifqi, Mas Edo, Mas Rizki, dan Mas Ikal). Terima kasih telah memberikan semangat, motivasi, dukungan, dan suka cita selama kegiatan magang di PT XYZ;
8. Teman-teman TIF 2013 seperjuangan (Amelia Fahmi, Bagus Aryo Pamungkas, Deste Mohammad Salim, Febbie Ramadhini, Fadillah Indra, Fildzah Adra Arifah, Fitriah Febriani, Gusti Maulana Arif, Iman Nurmansyah, Jimmy, Lilyani Barrung, Muhammad Khalish Ramadhansyah, Ridho Gilang Fiesta, Rizky Novriyedi Putra, Yusuf Arwadi). Terima kasih telah menjadi teman yang selalu memberikan semangat, motivasi, dukungan dan suka cita selama 4 tahun ini.
9. Teman-teman Universitas Bakrie Angkatan 2013 yang tidak bisa disebutkan satu per satu yang telah memberikan semangat selama penyusunan Tugas Akhir ini.
10. Kakak-kakak senior, adik-adik & Alumni TIF yang telah memberikan do'a, dukungan, dan semangat selama masa perkuliahan dan penyusunan Tugas Akhir.
11. Seluruh pihak yang terlibat dalam penyusunan Tugas Akhir ini yang tidak dapat penulis sebutkan satu persatu;

Dengan segala keterbatasan yang ada, penulis menyadari bahwa penyusunan tugas akhir ini masih jauh dari kesempurnaan. Untuk itu, saran dan kritik akan selalu diterima agar penulis dapat memperbaiki setiap kekurangan untuk kesempurnaan dimasa mendatang.

Akhirnya, penulis menyampaikan ucapan terima kasih dan semoga Allah SWT membalas segala kebaikan serta melimpahkan berkat dan rahmat-Nya kepada semua pihak yang telah membantu selama ini. Penulis berharap semoga Tugas Akhir ini berguna dan bermanfaat bagi kita semua.

Jakarta, 12 September 2017



Millah Fatimah

Universitas Bakrie

HALAMAN PERNYATAAN PERSUTUJUAN PUBLIKASI

Sebagai *civitas* akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Millah Fatimah
NIM : 1132001019
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Jenis Tugas Akhir : Implementasi Algoritma

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Prediksi *Cost Per Action* (CPA) Pada Iklan Facebook PT XYZ Menggunakan Algoritma Random Forest

Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pngkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 14 September 2017

Yang menyatakan



Millah Fatimah

v

Prediksi *Cost Per Action* (CPA) Pada Iklan Facebook PT XYZ Menggunakan Algoritma Random Forest

Millah Fatimah

ABSTRAK

Penelitian ini membahas tentang implementasi algoritma untuk dapat memprediksi *Cost Per Action* (CPA) yang ada pada iklan Facebook PT XYZ. Metode yang digunakan dalam penelitian untuk memprediksi CPA adalah dengan menerapkan algoritma Random Forest. Sistem prediksi ini bertujuan membantu praktisi *digital marketing* di divisi *Growth & Digital Marketing* PT XYZ agar dapat membantu menganalisa apakah iklan yang ada termasuk ke dalam kelompok CPA tinggi atau CPA rendah, agar pengeluaran budget perusahaan untuk iklan dapat diminimalisir. Sistem akan dibangun dengan menggunakan bahasa pemrograman Python. Hasil dari sistem prediksi CPA lalu akan dihitung nilai akurasi dengan menggunakan *Confusion Matrix*, perhitungan *Recall*, *Precision*, *Weighted Accuracy*, *Specificity*, dan *ROC Curve* serta akan dilihat fitur-fitur apa sajakah yang berpengaruh dalam sistem prediksi dengan menggunakan *Feature Importance*. Hasil dari penelitian ini adalah diperoleh fitur-fitur yang paling berpengaruh dalam proses prediksi menggunakan algoritma Random Forest menggunakan data *training*, data *testing* pada data model I, data model II, dan data model III yakni fitur *reach*, *amount_spent*, *budget type*. Selain itu juga diperoleh tingkat performa prediksi yang rendah pada data model I, yakni rata-rata 58% untuk data *training* & data *testing*, serta sangat baik pada data model II dan data model III dengan rata-rata akurasi 92% dan 94%.

Kata kunci: *Supervised Learning*, Random Forest, Prediksi Data, Klasifikasi Data

**Prediksi *Cost Per Action* (CPA) Pada Iklan Facebook PT XYZ Menggunakan Algoritma
Random Forest**

Millah Fatimah

ABSTRACT

This research discusses algorithm implementation to predict Cost per Action (CPA) in Facebook Ads PT XYZ. The method that will be used in this research to predict CPA is by implementing Random Forest Algorithm. The system aims to assist digital marketing practitioners in PT XYZ Growth & Digital Marketing team in order to analyzing whether the ad will be classified in High CPA class or Low CPA class, thus it can help company in order to minimize the company's online advertising cost. The system will be developed by using Python programming language. The result of this prediction system will be evaluated through Confusion Matrix, Recall, Precision, Weighted Accuracy, Specificity measure, and ROC Curve, along with Feature Importance graph to describe the features that have high influence in this prediction system. The result of this research are getting the most influence features involved in the prediction system, which are reach, amount_spent, budget type. The prediction system also have pretty nice results, proved with a nice value of accuracy from training & testing data which is 58% in average, and very good value of accuracy from model II and model III data which is 92% and 94% in average.

Keyword: *Supervised Learning, Random Forest, Data Prediction, Data Classification*

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
UNGKAPAN TERIMA KASIH.....	4
HALAMAN PERNYATAAN PERSUTUJUAN PUBLIKASI	Error! Bookmark not defined.
ABSTRAK.....	7
ABSTRACT.....	8
DAFTAR ISI.....	9
DAFTAR GAMBAR.....	11
DAFTAR TABEL.....	13
DAFTAR PERSAMAAN	15
DAFTAR ALGORITMA.....	16
DAFTAR LAMPIRAN.....	17
BAB I Pendahuluan	Error! Bookmark not defined.
1.1 Latar Belakang Masalah.....	Error! Bookmark not defined.
1.2 Perumusan Masalah	Error! Bookmark not defined.
1.3 Batasan Masalah	Error! Bookmark not defined.
1.4 Tujuan Penelitian	Error! Bookmark not defined.
1.5 Manfaat Penelitian	Error! Bookmark not defined.
1.6 Sistematika Penulisan	Error! Bookmark not defined.
BAB II Tinjauan Pustaka	Error! Bookmark not defined.
2.1 Penelitian Terkait	Error! Bookmark not defined.
2.2 Facebook Ads.....	Error! Bookmark not defined.
2.2.1 <i>Cost Per Action</i> (CPA).....	Error! Bookmark not defined.
2.2.2 <i>Ads Objective</i>	Error! Bookmark not defined.
2.3 Metode Prediksi CPA	Error! Bookmark not defined.
2.3.1 <i>Data Mining</i>	Error! Bookmark not defined.
2.3.2 Metode Random Forest Sebagai Metode Untuk Prediksi CPA.....	Error! Bookmark not defined.
2.3.3 Pengukuran Performa dan Akurasi Prediksi	Error! Bookmark not defined.
BAB III Metodologi Penelitian.....	Error! Bookmark not defined.
3.1 Studi Literatur	Error! Bookmark not defined.
3.2 Studi Prosedur Yang Sedang Berjalan & Identifikasi Masalah.....	Error! Bookmark not defined.

3.3	Objek Penelitian.....	Error! Bookmark not defined.
3.4	Pengumpulan Data	Error! Bookmark not defined.
3.4.1	Data <i>Training</i>	Error! Bookmark not defined.
3.4.2	Data Uji / <i>Testing</i>	Error! Bookmark not defined.
3.5	Pemilihan Fitur dengan Random Forest.....	Error! Bookmark not defined.
3.6	Prediksi dengan metode Random Forest.....	Error! Bookmark not defined.
3.7	Pengukuran Akurasi Prediksi dan Performa	Error! Bookmark not defined.
3.8	Pembahasan.....	Error! Bookmark not defined.
3.9	Penyusunan Laporan Penelitian	Error! Bookmark not defined.
BAB IV IMPLEMENTASI & PEMBAHASAN		Error! Bookmark not defined.
4.1	Data Masukan	Error! Bookmark not defined.
4.2	Persiapan Data	Error! Bookmark not defined.
4.3	Pembentukan Pohon Keputusan C4.5	Error! Bookmark not defined.
4.5	Uji Akurasi Prediksi & <i>Feature Impotance</i> Random Forest.....	Error! Bookmark not defined.
BAB V		Error! Bookmark not defined.
SIMPULAN DAN SARAN		Error! Bookmark not defined.
5.1	Simpulan	Error! Bookmark not defined.
5.2	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA		Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar 2.1 Langkah-langkah pada Jaringan CPA.....11

Gambar 2.2 Proses *Knowledge Discovery Database*..... 17

Gambar 2.3 Struktur Simpul Pohon Keputusan..... 19

Gambar 2.4 *Flowchart* Algoritma Random Forest..... 22

Gambar 2.5 Kurva *Receiving Operator Characteristics* (ROC).....27

Gambar 3.1 Tahapan Penelitian 28

Gambar 3.2 Struktur Organisasi Divisi *Growth & Digital Marketing* PT XYZ..... 30

Gambar 3.3 Tahapan Pembentukan Dataset Menggunakan Random Forest dan Implementasi Random Forest Untuk prediksi Serta Pengukuran Akurasi Prediksi dan Performa 33

Gambar 3.4 Proses Prediksi Algoritma Random Forest35

Gambar 4.1 *Script* Perhitungan *Entropy* Awal.....42

Gambar 4.2 *Script* Perhitungan *Entropy* Akhir & *Information Gain*..... 42

Gambar 4.3 *Script* Pembentukan Random Forest..... 47

Gambar 4.4 *Script* Prediksi Random Forest..... 48

Gambar 4.5 *Script Confusion Matrix*..... 48

Gambar 4.6 *Script Accuracy*..... 49

Gambar 4.7 *Script Precision*..... 50

Gambar 4.8 *Script Recall*..... 50

Gambar 4.9 *Script Specificity*..... 51

Gambar 4.10 *Script Weighted Accuracy*..... 51

Gambar 4.11 Kurva ROC data training model I..... 53

Gambar 4.12 Grafik *Feature Importance* data training model I..... 53

Gambar 4.13 Kurva ROC data testing model I..... 54

Gambar 4.14 Grafik *Feature Importance* data testing model I..... 57

Gambar 4.15 Kurva ROC data training model II.....61

Gambar 4.16 Grafik *Feature Importance* data training model II.....61

Gambar 4.17 Kurva ROC data data *testing* model II.....64

Gambar 4.18 Grafik *Feature Importance* data testing model II.....68

Gambar 4.19 Kurva ROC data *testing* model III.....64

Gambar 4.20 Grafik *Feature Importance* data *testing* model III.....65

Gambar 4.21 Kurva ROC data *training* model III.....71

Gambar 4.22 Grafik *Feature Importance* data *training* model III.....71

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait.....	11
Tabel 2.2 <i>Confusion matrix</i> multi kelas.....	27
Tabel 3.1 <i>Role</i> dan Deskripsi Kerja Divisi <i>Growth & Digital Marketing</i> PT XYZ.....	31
Tabel 3.2 Contoh Data Pengelompokkan CPA	33
Tabel 3.3 Alokasi Waktu Penelitian	39
Tabel 4.1 Fitur data performa iklan model I.....	40
Tabel 4.2 Data mentah CPA.....	41
Tabel 4.3 Data CPA Terklasifikasi.....	41
Tabel 4.4 Fitur <i>result_indicator</i>	42
Tabel 4.5 Fitur <i>year</i>	42
Tabel 4.6 Fitur label (CPA)	42
Tabel 4.7 Nilai Entropy & Information Gain Putaran ke-1.....	43
Tabel 4.8 Nilai Entropy & Information Gain Putaran ke-2.....	44
Tabel 4.9 Nilai Entropy & Information Gain Putaran ke-3.....	45
Tabel 4.10 Nilai Entropy & Information Gain Putaran ke-4.....	45
Tabel 4.11 Nilai Entropy & Information Gain Putaran ke-5.....	46
Tabel 4.12 Nilai Entropy & Information Gain Putaran ke-6.....	46
Tabel 4.13 Nilai Entropy & Information Gain Putaran ke-7.....	47
Tabel 4.14 <i>Confusion Matrix</i> data training model I.....	51
Tabel 4.15 Nilai perhitungan akurasi prediksi data training model I.....	53
Tabel 4.16 Urutan <i>Feature Importance</i> data training model I.....	55
Tabel 4.17 <i>Confusion Matrix</i> data testing model I.....	55
Tabel 4.18 Nilai perhitungan akurasi prediksi data testing model I.....	56
Tabel 4.19 Urutan <i>Feature Importance</i> data testing model I.....	58
Tabel 4.20 Fitur data performa iklan model II.....	59
Tabel 4.21 Fitur <i>budget_type</i>	59
Tabel 4.22 <i>Confusion Matrix</i> data testing model II.....	59

Tabel 4.23 Nilai perhitungan akurasi prediksi data testing model II.....	60
Tabel 4.24 Urutan Feature Importance data testing model II.....	62
Tabel 4.25 Confusion Matrix data training model II.....	62
Tabel 4.26 Nilai perhitungan akurasi prediksi data training model II.....	63
Tabel 4.27 Urutan Feature Importance data training model II.....	65
Tabel 4.28 Fitur data performa iklan model III.....	65
Tabel 4.29 Confusion Matrix data testing model III.....	66
Tabel 4.30 Nilai perhitungan akurasi prediksi data testing model III.....	67
Tabel 4.31 Urutan Feature Importance data testing model III.....	69
Tabel 4.32 Confusion Matrix data training model III.....	69
Tabel 4.33 Nilai perhitungan akurasi prediksi data training model III.....	70
Tabel 4.34 Urutan Feature Importance data training model III.....	72
Tabel 4.35 Ringkasan Metriks Akurasi Prediksi.....	72

DAFTAR PERSAMAAN

Persamaan (2.1)	Indeks Gini
Persamaan (2.2)	Indeks Informasi
Persamaan (2.3)	Indeks Twoing
Persamaan (2.4)	Indeks Entropi
Persamaan (2.5)	<i>Goodness of Split</i>
Persamaan (2.6)	Penandaan Label Kelas
Persamaan (2.7)	<i>Entropy</i>
Persamaan (2.8)	<i>Information Gain</i>
Persamaan (2.9)	<i>Specificity</i>
Persamaan (2.10)	<i>Recall</i>
Persamaan (2.11)	<i>Precision</i>
Persamaan (2.12)	<i>Accuracy</i>
Persamaan (2.13)	<i>Weighted Accuracy</i>

DAFTAR ALGORITMA

Algoritma 3.1 *Pseudocode* Random Forest

DAFTAR LAMPIRAN

Lampiran 1 Transkrip Wawancara.....**Error! Bookmark not defined.**
Lampiran 2 Dataset Training**Error! Bookmark not defined.**
Lampiran 3 Data Testing**Error! Bookmark not defined.**
Lampiran 4 Data Training Model II.....**Error! Bookmark not defined.**
Lampiran 5 Data Testing Model II**Error! Bookmark not defined.**
Lampiran 6 Data Training Model III**Error! Bookmark not defined.**
Lampiran 7 Data Testing Model III**Error! Bookmark not defined.**
Lampiran 8 Data Mentah Performa Iklan**Error! Bookmark not defined.**
Lampiran 9 Script Pembentukan Pohon Keputusan C4.5**Error! Bookmark not defined.**
Lampiran 10 Script Random Forest**Error! Bookmark not defined.**
Lampiran 11 Script Confusion Matrix & Uji Akurasi Prediksi ..**Error! Bookmark not defined.**
Lampiran 12 Script ROC Curve.....**Error! Bookmark not defined.**
Lampiran 13 Script Feature Importance**Error! Bookmark not defined.**