

# Estimasi Data Insight Social Media Ads Menggunakan Neural Network, Linear Regression dan Deep Learning

Zurni Laila, Nuri Cahyono\*

Fakultas Ilmu Komputer, S1 Informatika, Universitas Amikom Yogyakarta, Sleman, Indonesia

Email: <sup>1</sup>zurni.laila@students.amikom.ac.id, <sup>2</sup>\*nuricahyono@amikom.ac.id

Email Penulis Korespondensi: nuricahyono@amikom.ac.id

Submitted: 01/01/2023; Accepted: 31/03/2023; Published: 31/03/2023

**Abstrak**—PT IlmuKomputerCom Braindevs merupakan perusahaan profesional training yang menjual produk berupa jasa pelatihan (course). Pada perusahaan jasa seperti PT IlmuKomputerCom Braindevs Sistemanya terselenggaranya course dengan peminat yang tinggi adalah kunci untuk peningkatan profit perusahaan. Divisi marketing merupakan divisi yang sangat penting dalam rangka terselenggaranya suatu course/training. Untuk menyelenggarakan dibutuhkan target peserta dalam menyelenggarakan training. Selain itu PT IlmuKomputerCom Braindevs juga perlu mengestimasi biaya iklan dan durasi iklan dalam promosi training yang akan diselenggarakan. Untuk menganalisis divisi marketing menggunakan teknik data mining dan metode Cross-industry standard process for data mining (CRISP-DM) untuk mendapatkan estimasi yang diinginkan. Maka untuk mendapatkan hasil akhir nilai estimasi peserta, durasi iklan dan biaya iklan dibutuhkan algoritma yang memiliki akurasi yang paling dengan acuan dari hasil perbandingan algoritma dengan melihat nilai dari RMSE (Root Mean Square Error). Semakin nilai yang dihasilkan mendekati nilai 0 maka nilai akurasi RMSE (Root Mean Square Error) estimasi yang dihasilkan akan semakin baik.

**Kata Kunci:** Data Mining; CRISP-DM; RMSE; Estimasi

**Abstract**—PT IlmuKomputerCom Braindevs is a professional training company that sells products in the form of training services (courses). In service companies such as PT IlmuKomputerCom Braindevs Sistemanya, holding courses with high demand is the key to increasing company profits. The marketing division is a very important division in the context of holding a course/training. To manage the target participants needed in organizing training. In addition, PT IlmuKomputerCom Braindevs also needs to estimate advertising costs and ad duration in the training promotions that will be held. To analyze the marketing division using data mining techniques and the Cross-industry standard process for data mining (CRISP-DM) method to obtain the desired estimate. So to get the final result of the participant's estimated value, ad duration and ad cost, an algorithm that has the most accurate accuracy is needed according to the reference from the results of the comparison of algorithms by looking at the value of RMSE (Root Mean Square Error). The closer the resulting value is to 0, the better the estimated accuracy of the RMSE (Root Mean Square Error) estimate will be.

**Keywords:** Data Mining; CRISP-DM; RMSE; Estimation.

## 1. PENDAHULUAN

Pada perusahaan profesional pelatihan seperti PT IlmuKomputerCom Braindevs Sistemanya, terselenggaranya kursus dengan peminat yang tinggi adalah kunci untuk peningkatan keuntungan perusahaan. Setiap produk kursus memiliki daya jual yang berbeda. Daya jual kursus diukur dengan berapa banyak kursus yang sudah berhasil diselenggarakan. Untuk itu, perusahaan harus menargetkan jumlah peserta pada setiap pelatihan yang akan dilaksanakan. Peserta pelatihan merupakan kunci terlaksananya sebuah pelatihan.

Produk training pada perusahaan ini meliputi training dibidang teknologi seperti data science, Software engineer, Enterprise Architecture serta produk course dengan topik teknologi terupdate. Untuk itu setiap ada launching course terbaru dengan topik trend teknologi maka dibutuhkan analisis pasar sehingga product bisa diterima customer maupun peserta yang berlangganan mengikuti course pada perusahaan profesional training.

Melakukan analisis pada data peserta training menggunakan teknik data mining untuk melakukan estimasi jumlah peserta yang ideal dalam menyelenggarakan sebuah training sehingga akan dijadikan acuan jumlah peserta dalam menyelenggarakan training course selanjutnya. Perusahaan kesulitan untuk menentukan pengeluaran biaya iklan dalam melakukan promosi berbayar. Maka perlu dilakukan analisis prediksi biaya iklan agar lebih mudah dalam mengatur rancangan keuangan perusahaan.

Untuk meningkatkan peserta pelatihan/training suatu perusahaan maka perlu melakukan promosi atau iklan training agar informasi training yang akan dilaksanakan bisa menjangkau lebih banyak audience sehingga peserta yang ikut bisa memenuhi target. Salah satu cara untuk memperluas informasi ialah dengan menggunakan iklan di sosial media seperti iklan pada facebook, instagram dan sosial media lainnya. Namun untuk mendapatkan jangkauan luas pada sosial media tidak cukup hanya dengan memposting sebuah informasi maka diperlukannya iklan pendukung seperti menggunakan iklan ads berbayar pada sosial media tertentu.

Menerapkan Teknik data mining yaitu estimasi dalam pemrosesan setiap data. Menggunakan Metode estimasi data mining untuk keperluan perusahaan yaitu untuk mengetahui jumlah peserta yang ideal, Estimasi biaya iklan yang perlu dikeluarkan dalam melakukan promosi berbayar dan juga estimasi durasi iklan promosi berbayar. Dengan adanya perkiraan dan estimasi yang dilakukan perusahaan bisa meningkatkan kualitas perusahaan dalam menyelenggarakan training dan bisa meningkatkan profit pada perusahaan. Adapun beberapa algoritma yang akan dipakai untuk melakukan proses estimasi adalah Algoritma Neural Network (NN), Generalize

linear Model (GLM), Linear regression, Support Vector Machine (SVM) dan Deep Learning. Penelitian ini juga bertujuan untuk melakukan perbandingan dan menentukan algoritma terbaik dalam memprediksi suatu permasalahan. Membandingkan beberapa algoritma untuk mendapatkan hasil terbaik dalam melakukan estimasi yang dibutuhkan perusahaan. Dalam proses modeling penelitian ini akan menggunakan teknik Cross-Industry Standard Process for Data Mining (CRISP-DM) sebagai konsep yang sering digunakan dalam proses pemecahan permasalahan pada perusahaan industry untuk mengambil keputusan dari permasalahan.

Dengan adanya estimasi jumlah peserta yang ideal, Estimasi biaya iklan yang perlu dikeluarkan dalam melakukan promosi berbayar dan juga estimasi durasi iklan promosi berbayar bisa meningkatkan profit perusahaan karena sudah mengacu pada hasil analisis menggunakan data mining. Dengan mengatur budget pengeluaran lebih terencana dan juga Training Course yang direncanakan bisa terlaksana dengan maksimal karena dibantu dengan durasi promosi iklan yang maksimal. Pemrosesan Menggunakan data yang diambil dari divisi marketing pada perusahaan PT IlmuKomputerCom Braindevs Sistem.

Penelitian ini bertujuan untuk melakukan prediksi jumlah peserta yang ideal dalam menyelenggarakan pelatihan/training pada PT IlmuKomputerCom Braindevs sehingga training bisa berjalan dengan efektif. Selain untuk memprediksi peserta traini tujuannya ialah untuk mengestimasi biaya iklan dan durasi iklan pada setiap informasi training sehingga bisa menjangkau peserta yang sudah ditargetkan. Menerapkan Teknik data mining yaitu estimasi dalam pemrosesan setiap data. Adapun algoritma yang akan diimplementasikan dalam proses modeling ini adalah algoritma Generalize linear Model (GLM), Linear regression, Neural Network, Support Vector Machine (SVM) dan Deep Learning. Penelitian ini juga bertujuan untuk menentukan algoritma terbaik dalam memprediksi suatu permasalahan. Menggunakan konsep Cross-Industry Standard Process for Data Mining (CRISP-DM) sebagai teknik yang banyak digunakan dan diimplementasikan dalam perusahaan industry untuk memecahkan masalah dan mengambil keputusan dari suatu permasalahan.

Data mining ialah proses ekstraksi untuk mencari model dan informasi pada suatu data menggunakan metode tertentu pada data mining. Tujuan dari data mining ialah mencari insight dan indikasi yang bermanfaat yang bisa digunakan dalam suatu perusahaan maupun kehidupan sehari-hari. Pemilihan algoritma yang tepat sangat mempengaruhi hasil proses pemodelan. Metode data mining membuktikan dalam mengatasi dan memecahkan masalah darurat baik dibidang bisnis dan bidang lainnya[1]. Adapun teknik dalam data mining sangat bervariasi diantaranya yang sering digunakan adalah prediksi (Prediction), Estimasi (Estimation), Klasifikasi (classification), Klusterisasi (Clustering), dan Asosiasi (Association). Berikut teknik data mining berdasarkan tugas yang dapat dilakukan[2]. Neural Network adalah implementasi ilmu yang diadaptasi dari kemampuan otak manusia yang bisa memberikan rangsangan dan stimulus dalam melakukan proses sehingga bisa menghasilkan output yang diinginkan. Neural networks atau artificial neural networks telah menjadi alat standar yang penting untuk data mining dan banyak digunakan untuk tugas-tugas data mining seperti pattern classification, time series analysis, prediction dan clustering. Dalam praktek neural network digunakan untuk melakukan pemodelan kompleks antara relasi input dan output dalam mencari suatu model dalam data. Neural network banyak digunakan melakukan pemecahan masalah dan mengambil keputusan berdasarkan data yang diproses. Kelebihan dari neural network ialah sangat bagus untuk proses pengolahan pada data noise. Sedangkan kelemahan algoritma ini sulit dimengerti oleh pemula. Linear Regression adalah metode untuk melakukan pemodelan antara hubungan satu variable dependen dan satu variable independent. Linear Regression atau Regresi Linear menemukan garis terbaik dari 2 atribut atau variable, sehingga salah satu atribut tersebut dapat memprediksi atribut yang lain. Model regresi dan log-linear dapat digunakan untuk memperkirakan data yang diberikan[3]. Proses model regresi linear bertujuan untuk menemukan garis yang paling baik antar atribut dan variable. Regresi Linear juga berfungsi untuk memprediksi nilai yang belum diketahui menggunakan nilai data yang memiliki keterkaitan dan nilainya sudah diketahui. Support Vector Machine atau biasa disingkat dengan SVM ialah algoritma yang termasuk pada supervised learning yang diimplementasikan pada pemrosesan klasifikasi, prediksi maupun regresi pada suatu object permasalahan. Konsep SVM ialah berdasarkan Structural Risk Minimization (SRM), minimalisasi error ketika training dan memaksimalkan margin[4] yang ditujukan untuk melakukan pengolahan menjadi dua kelas berbeda atau biasa disebut Hyperplane[5]. Generalized Linear Model merupakan model sambungan dari linear regression dengan mengasumsikan distribusi spesifik dari variabel. Metode GLM adalah Kuantifikasi Faktor independen dengan faktor (variabel respon) Binding (variabel prediktor), penggunaan Cara ini untuk menjelaskan perubahannya. Jika ada, nilai faktor besar Perubahan faktor independen. Pengujian prediksi algoritma Generalized Linear didasarkan dengan nilai mean square error (MSE), root mean square error (RMSE), dan mean absolute persentase error (MAPE)[6]. Cross validation ialah metode yang digunakan dalam tahap evaluasi kinerja pada suatu algoritma atau model yang sudah dilakukan pemrosesan dengan menggeneralisasi kumpulan pada data independen. Pada tahap cross validation model dianalisis dan di evaluasi berdasarkan parameter setiap algoritma. Teknik ini utamanya digunakan untuk melakukan prediksi model dan memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan dalam praktiknya[7]. Root Mean Square Error (RMSE) ialah sebuah metode untuk mengukur perbedaan nilai evaluasi dari model estimasi. Root Mean Square Error (RMSE) diukur berdasarkan nilai index error pada suatu model. Acuan pengambilan nilai pada Root Mean Square Error (RMSE) ialah semakin nilai mendekati 0 maka akurasi model semakin baik. Root means square error juga dikenal sebagai kesalahan standar asumsi atau perkiraan yang dijadikan standar acuan ketepatan estimasi[3]. Evaluasi RMSI ini sering digunakan dan diterapkan untuk melihat efektifitas pada model estimasi yang sudah diproses. RapidMiner

ialah suatu software atau perangkat lunak yang didedikasikan untuk melakukan proses data mining. RapidMiner memberikan kemudahan untuk melakukan analisis data mining seperti melakukan prediksi, estimasi, clusterisasi, klasifikasi dan asosiasi. Pada perangkat lunak RapidMiner juga dilengkapi kemudahan untuk melakukan analisis text meaning, pembelajaran machine learning, predictive analytics dan deep learning. RapidMiner merupakan sebuah solusi untuk melakukan analisis bisnis komersial, Proses penelitian, pendidikan dan pembuatan prototype dengan cepat. Aplikasi ini sangat mendukung semua proses pengolahan machine learning dan data mining seperti persiapan data, visualisasi data, validasi data dan optimalisasi pada suatu data[7].

Beberapa penelitian terdahulu yang sudah dilakukan oleh beberapa penulis diantaranya oleh Fitriyani & Rangga Sanjaya, (2018) yang berjudul "KOMPARASI ALGORITMA LR, K-NN DAN SVM UNTUK ESTIMASI AREA KEBAKARAN HUTAN" yang membahas tentang dampak kebakaran hutan dan pengukuran tingkat radiasi api pada area terbakar menggunakan algoritma LR (Linear Regression), K-NN (K-Nearest Neighbor) dan SVM (Support Vector Machine). Dalam Penelitian ini menggunakan perbandingan tiga algoritma yaitu algoritma Linear Regression (LR), K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM), Dari perbandingan dari tiga algoritma ini dilakukan evaluasi untuk menghasilkan algoritma yang memiliki akurasi terbaik untuk melakukan pemodelan pada estimasi area kebakaran hutan. Setelah melakukan evaluasi perbandingan dihasilkan bahwa K-Nearest Neighbor merupakan algoritma yang sangat populer dan banyak digunakan, Sedangkan pada algoritma SVM termasuk algoritma yang cukup baik untuk mengatasi permasalahan klasifikasi dan regresi baik linier maupun nonlinier. Penelitian ini mengambil data dari UCI Repository untuk melakukan proses pemodelan. Dari proses pemodelan evaluasi menghasilkan algoritma terbaik untuk melakukan estimasi kebakaran hutan ialah Linear Regression (LR) dengan hasil RMSE sebesar 1.496, Kemudian algoritma terbaik kedua ialah algoritma K-Nearest Neighbor (K-NN) dengan hasil nilai evaluasi sebesar 1.981 dan Algoritma terbaik ketiga dari hasil uji validasi nilai RMSE ialah algoritma Support Vector Machine (SVM) dengan nilai 3.088[8]. Dari penelitian ini maka penulis menggunakan beberapa algoritma seperti neural network, linear regression dan Generalized Linear Models. Perbandingan penelitian ini dengan penelitian sebelumnya dari hasil perbandingan algoritma mendapatkan algoritma yang berbeda. Pada penelitian sebelumnya menghasilkan algoritma Linear regression sebagai algoritma terbaik. Sedangkan pada penelitian ini membandingkan beberapa algoritma dengan menggunakan objek data yang berbeda dan menghasilkan algoritma terbaik yang berbeda sesuai object estimasi yang diinginkan.

Kemudian penelitian yang dilakukan oleh Tesa Nur Padilah & Riza Ibnu Adam, (2019) melakukan penelitian dengan judul "ANALISIS REGRESI LINIER BERGANDA DALAM ESTIMASI PRODUKTIVITAS TANAMAN PADI DI KABUPATEN KARAWANG" yang membahas tentang estimasi regresi linier berganda dalam menganalisis produktivitas pertanian. Pada penelitian ini menggunakan variabel terikat pada 57 data pada padi (kw/ha). Variabel bebas yang digunakan pada penelitian ini meliputi produksi, luas panen, luas tanaman, rata-rata curah hujan dan rata-rata hari hujan. Pada data tersebut dilakukan proses pemodelan menggunakan algoritma regresi linier berganda sehingga mendapatkan hasil yang ingin di estimasi. Adapun hasil dari penelitian ini adalah hari hujan. Sedangkan sisanya 19,54% dapat dijelaskan oleh faktor-faktor lain yang tidak diteliti dalam penelitian ini. Variabel-variabel yang mempengaruhi peningkatan jumlah produktivitas padi yaitu variabel produksi dan curah hujan, sedangkan variabel-variabel yang mempengaruhi penurunan jumlah produktivitas yaitu variabel luas panen, luas tanam, dan hari hujan. Rata-rata kesalahan relatif regresi yang diperoleh yaitu 0,04642 atau 4,642% [9]. Pada penelitian ini fokus menggunakan satu algoritma untuk melakukan estimasi pada produktivitas tanaman padi yang menghasilkan nilai relatif kesalahan regresi sebesar 4,642%. Sedangkan penelitian yang dilakukan oleh penulis ialah penelitian perbandingan beberapa algoritma untuk melakukan estimasi. Menggunakan algoritma terbaik untuk melakukan modeling. Algoritma yang dianggap paling baik pada penelitian yang dilakukan penulis ialah algoritma neural network, linear regression dan Generalized Linear Models.

Penelitian selanjutnya dilakukan oleh Fitria Habibatul Hamdanah & Devi Fitriana (2021) dengan judul "ANALISIS PERFORMANS ALGORITMA LINEAR REGRESSION DENGAN GENERALIZED LINEAR MODEL UNTUK PREDIKSI PENJUALAN PADA USAHA MIKRO, KECIL, DAN MENENGAH". Pada penelitian dilakukan perbandingan algoritma yaitu Linear Regression dan Generalized Linear Model. Untuk melakukan perbandingan antara keduanya dilakukan uji validasi dengan acuan terhadap nilai RSME (root mean square error). Dari hasil nilai RSME (root mean square error) semakin nilai yang dihasilkan mendekati 0 maka semakin akurat hasil prediksi yang dilakukan. Data yang digunakan dalam proses ini ialah data penjualan UMKM yang diambil dari aplikasi PINTER USAHA untuk melakukan proses modeling perbandingan algoritma Linear Regression dan Generalized Linear Model. Hasil dari proses modeling menggunakan uji validation dapat disimpulkan bahwa algoritma Linear regression lebih baik dari pada Generalized Linear Model dengan nilai RMSE pada algoritma Linear ialah sebesar 1,983 sedangkan nilai hasil pada Generalized Linear Model ialah sebesar 4,827. Penelitian ini membuktikan bahwa dengan menggunakan algoritma Linear Regression dan Generalized Linear Model dapat memberikan prediksi terhadap penjualan pada UMKM di PINTER USAHA sehingga menghasilkan prediksi penjualan barang untuk tahun berikutnya[10]. Perbandingan penelitian ini ialah membandingkan algoritma linear regression dan generalized linear model dan mendapatkan nilai dari validasi sebesar 1,983 dan 4,827 sehingga algoritma keduanya di prediksi bagus untuk penelitian selanjutnya. Sedangkan pada penelitian yang dilakukan oleh penulis ialah perbandingan algoritma dan menjadikan proses modeling pada

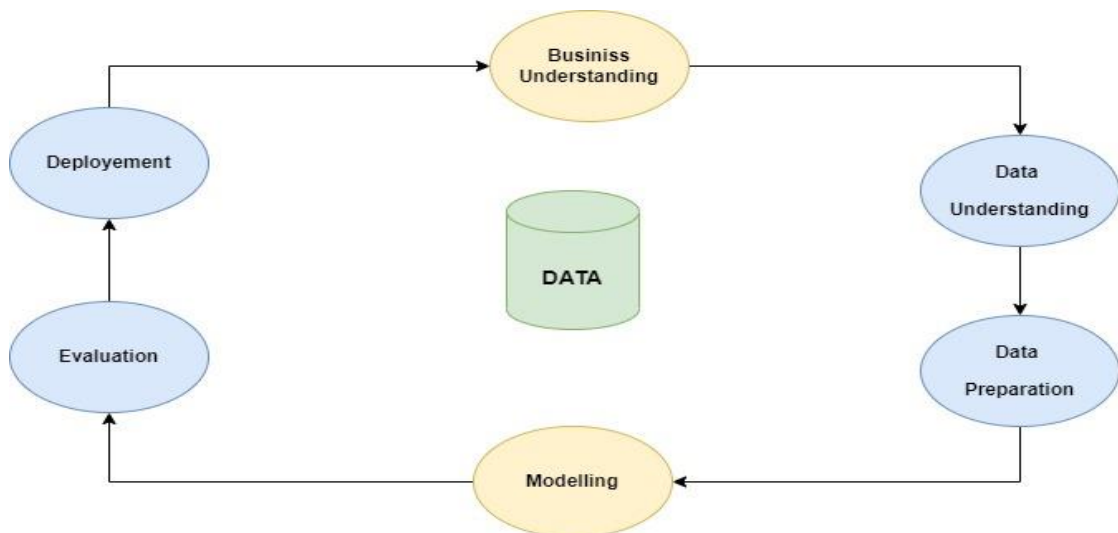
algoritma yang menghasilkan akurasi yang tinggi diukur dari uji validasi berdasarkan nilai root means square error. Algoritma yang paling baik digunakan untuk pemodelan estimasi setiap objek data.

Dan selanjutnya penelitian yang dilakukan oleh Benaya Pradito dan Dini Silvi Purnia,(2020) melakukan penelitian untuk melakukan peramalan pada mata uang kurs dengan judul penelitian"KOMPARASI ALGORITMA LINEAR REGRESSION DAN NEURAL NETWORK UNTUKMEMPREDIKSI NILAI KURS MATA UANG". Pada penelitian ini berfokus untuk melakukan prediksi nilai mata uang kurs. Selain itu penelitian ini membandingkan dua algoritma yaitu algoritma linear regression dan neural network. Pada hasil penelitian ini meghasilkan bahwa algoritma dengan prediksi terbaik adalah linear regression dengan RMSE Jual 28.012 +/- 5.626, Absolute Error 21.444 +/- 4.095, Relative Error 0.20% +/- 0.04% pada kurs jual dan RMSE Beli 27.556 +/- 5.893, Absolute Error 21.198 +/- 4.247, Relative Error 0.20% +/- 0.04%. Sehingga hasil prediksi yang didapatkan 10754.600 dan 10641.450 dari estimasi nilai mata uang kurs. Dan didapatkan pula linear regressioan memiliki akurasi yang lebih baik dibandingkan dengan nerval network[11]. Perbedaan dari penlitian ini adalah pada perbandingan algoritma. Dipenelitian ini membandingkan lima agoritma dari estimasi yaitu neural network, super vector machine, deep learning, linear regression dan Generalized Linear Models sehingga algoritma terbaik akan digunakan dalam melakukan estimas pada objek permasalahan.

## 2. METODOLOGI PENELITIAN

Proses pada penelitian Marketing untuk mendapatkan efisiensi promosi pada course yang akan diselenggarakan serta mendapatkan jumlah peserta yang ideal, Mengetahui durasi iklan dan estimasi biaya iklan pada course pada PT IlmuKomputerCom Braindevs Sistema menggunakan teknik Cross-industry standard process for data mining (CRISP-DM). Metode CRISP - DM memiliki beberapa tahapan proses yaitu Busniss Understanding, Data Understanding, Data Preparation, Modeling dan Evaluation[12]

Pada Penelitian ini menggunakan teknik CRISP-DM (Cross-Industry Standard Process for Data Mining) Yang sudah banyak diterapkan dan diimplementasikan dalam proses pemecahan masalah serta evaluasi bisnis pada perusahaan. CRISP-DM (Cross Industry Standard for Data mining) merupakan metode yang banyak digunakan oleh para ahli dengan menggunakan proses pemodelan data pada setiap pemodelan, Tujuan dari metode CRISP-DM ini yaitu untuk menganalisis dan menghasilkan suatu model atau pola yang menarik dan memiliki makna pada data yang digunakan dalam proses suatu model. Teknik CRISP-DM memiliki enam tahapan proses yaitu business understanding, data understanding, data preparation, modelling, evaluation, dan deployment. Berikut tahapan proses pada teknik CRISP-DM (Cross Industry Standard for Data mining)[13]:



**Gambar 1.** Tahapan Teknik Metode CRISP-DM

### 2.1 Business Understanding

Pada tahap Business Understanding ini melibatkan pemahaman tujuan dan persyaratan proyek dari perspektif bisnis, serta konversi tujuan tersebut ke dalam tujuan analisis data. Pada penelitian ini, tujuan bisnis adalah mengestimasi jumlah peserta, biaya promosi iklan, dan durasi promosi iklan berbayar dalam penyelenggaraan pelatihan.

### 2.2 Data Understanding

Tahap kedua setelah Business Understanding melibatkan pengumpulan data, evaluasi kualitas data, dan eksplorasi data awal. Dalam penelitian ini, data yang digunakan meliputi jumlah peserta, biaya promosi iklan, dan durasi promosi iklan berbayar.

### 2.3 Data Preparation

Data preparation merupakan proses ketiga dari teknik crisp-dm. Pada tahap ini, data dibersihkan, diproses, dan ditransformasi agar sesuai dengan kebutuhan analisis. Penelitian ini menggunakan data yang telah disiapkan dan diolah untuk memastikan kualitas data.

### 2.4 Modelling

Memasuki tahap modeling pada teknik CRISP-DM yang berfungsi untuk melakukan pemodelan pada data yang sudah melewati tahap preparation. Pemodelan melibatkan pemilihan teknik dan algoritma yang sesuai untuk mencapai tujuan bisnis, serta pembuatan model menggunakan data yang telah disiapkan. Pada penelitian ini, beberapa algoritma seperti Neural Network, Linear Regression, Support Vector Machine (SVM), Generalized Linear Model (GLM), dan Deep Learning digunakan dan dibandingkan menggunakan evaluasi cross-validation.

### 2.5 Evaluation

Di tahap evaluation untuk melakukan analisis pada hasil yang didapatkan ditahap modeling. Evaluasi melibatkan penilaian kinerja model yang telah dibuat, serta penentuan model mana yang paling sesuai dengan tujuan bisnis. Dalam penelitian ini, model dievaluasi menggunakan nilai Root Mean Square Error (RMSE) yang mendekati 0, dan model dengan nilai terbaik dipilih untuk digunakan dalam estimasi.

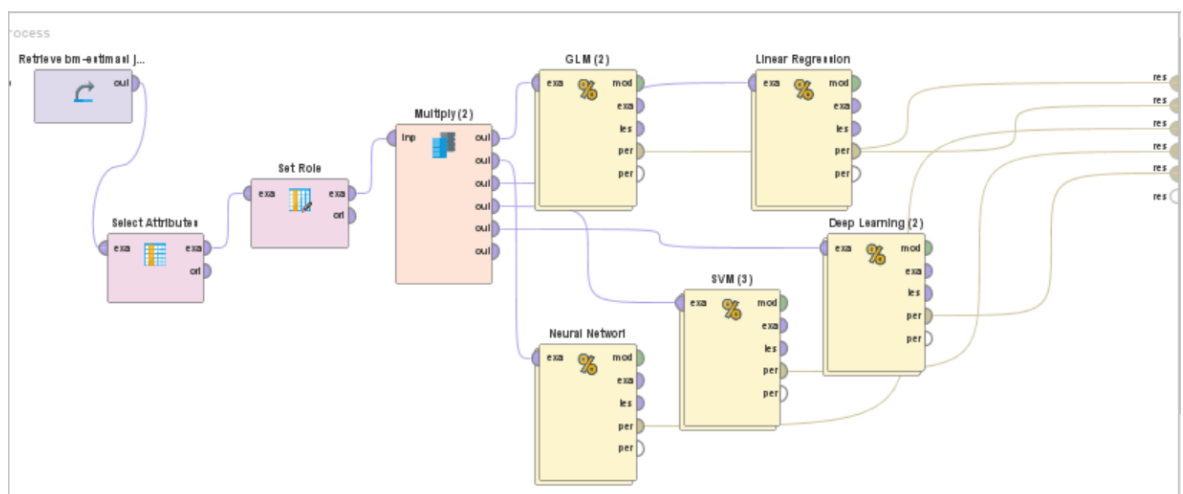
### 2.6 Deployment

Tahap deployment merupakan tahap akhir dari teknik crisp-dm. Ditahap ini untuk mengambil keputusan dari nilai hasil pemodelan yang sudah dilakukan serta melakukan implementasi pada perusahaan. Pada penelitian ini, model terbaik yang dihasilkan dari perbandingan algoritma digunakan untuk melakukan estimasi jumlah peserta, biaya promosi iklan, dan durasi promosi iklan berbayar.

## 3. HASIL DAN PEMBAHASAN

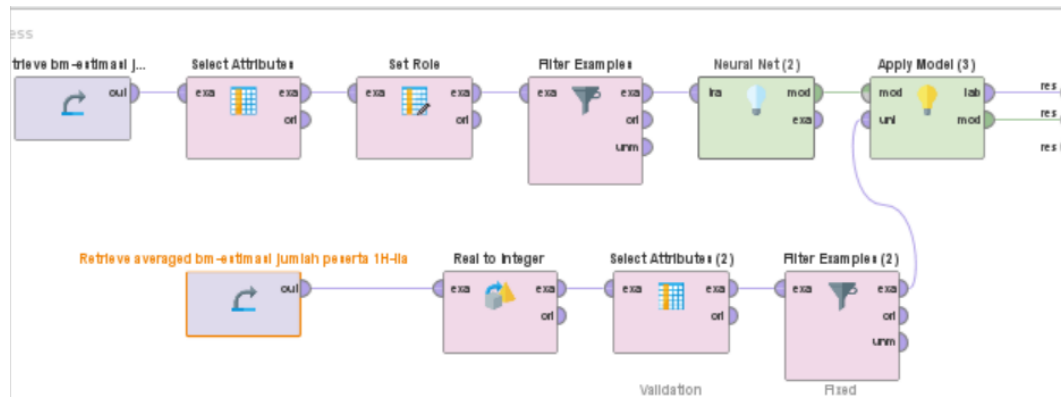
Pada tahap ini melakukan proses perbandingan algoritma menggunakan operator cross validation untuk mengetahui evaluasi pada suatu model dengan mengukur hasil dari nilai root mean square error yang mendekati 0, sehingga algoritma yang nilai rmse (Root mean Square Error) mendekati 0 akan digunakan untuk melakukan pemodelan selanjutnya. Teknik ini digunakan untuk proses prediksi model dan menganalisis akurasi setiap model[6]. Pada proses ini membanding beberapa algoritma menggunakan evaluasi cross validation, adapun algoritma yang dibandingkan adalah Neural Network, Linear regression, Support Vector Machine (SVM) dan , Generalized Linear Model (GLM).

Setiap perbandingan menggunakan Output yang berbeda sehingga nilai dan hasil perbandingan algoritma ditentukan oleh data yang akan digunakan serta hasil yang diinginkan. Setiap operator cross validation didalam nya terdapat algoritma yang akan di evaluasi serta dibandingkan nilainya .



**Gambar 2.** Proses Perbandingan Algoritma

Setelah melakukan proses modeling untuk membandingkan algoritma, selanjutnya melakukan modeling untuk mendapatkan hasil final dari estimasi yang diinginkan. Untuk mendapatkan hasil estimasi menggunakan data yang sudah disiapkan. Adapun algoritma yang mempunyai nilai terbaik untuk melakukan estimasi ialah Neural Network, Generalized Linear Model dan Linear Regression. Sehingga ketiga algoritma terbaik ini akan digunakan pada tahap pemodelan pengambilan keputusan. Berikut penjelasan perbandingan yang sudah diperoleh:



**Gambar 3.** Modeling Penentuan Estimasi

### 3.1 Estimasi Jumlah Peserta

Analisis Pengujian Perbandingan algoritma untuk menentukan Estimasi Jumlah peserta dengan mengukur performance menggunakan hitungan nilai hasil evaluasi pada RMSE (Root Mean Square Error). Pada setiap algoritma dilakukan pengujian menggunakan uji validasi menggunakan operator cross validation pada rapidminer. Adapun hasil Perbandingan nilai RMSE (Root Mean Square Error) pada beberapa proses algoritma sebagai berikut.

**Tabel 1.** Perbandingan Algoritma untuk melakukan estimasi peserta

Nama Algoritma	Nilai RMSE
Neural Network	79.243
Linear Regression	186.725
Support Vector Machine	59.624
Generalized Linear Model	49.984
Deep Learning	71.108

Setelah melakukan perbandingan algoritma maka algoritma yang hasil nilainya mendekati nilai 0 merupakan algoritma terbaik. Algoritma terbaik yang dihasilkan dalam perbandingan akan digunakan untuk melakukan estimasi. Adapun nilai evaluasi algoritma yang paling mendekati 0 ialah algoritma Generalized Linear Model (GLM) dengan nilai 49.984 sehingga GLM akan digunakan untuk melakukan estimasi jumlah peserta dalam penyelenggaraan training. Maka proses prediksi untuk menentukan Jumlah peserta yang ideal menyeleggarakan sebuah course. Hasil yang didapatkan dari modelling menggunakan Generalized Linear Model (GLM) Dengan menggunakan data rata – rata peserta, didapatkan bahwa jumlah peserta yang efektif dan ideal adalah sekitar 127 orang untuk menyeleggarakan training.

### 3.2 Estimasi Biaya Promosi Iklan

Melakukan perbandingan algoritma untuk estimasi biaya iklan estimasi yang meliputi beberapa algoritma diantaranya adalah Linear regression, Neural Network, Support Vector Machine, Generalized Model dan Deep Learning yang diuji menggunakan uji validasi cross validation. Hasil dari perbandinga uji validasi menentukan bahwa Neural Network merupakan algoritma terbaik dengan nilai 112929.701. Berikut hasil perbandingan algoritma dari hasil uji validasi estimasi biaya promosi iklan.

**Tabel 2.** Perbandingan Algoritma untuk melakukan estimasi biaya iklan

Nama Algoritma	Nilai RMSE
Neural Network	112929.701
Linear Regression	130128.142
Support Vector Machine	136001.433
Generalized Linear Model	130086.713
Deep Learning	125675.038

Setelah dihasilkan algoritma terbaik dari pengujian validasi untuk melakukan estimasi pada biaya promosi iklan maka dilakukan modelling menggunakan algoritma Neural Network dalam menentukan biaya promosi iklan sehingga bisa menjangkau banyak audience dan mendapatkan banyak peserta pada training yang akan dilaksanakan. Adapun hasil yang didapatkan setelah melakukan modeling menggunakan neural network maka dihasilkan bahwa biaya yang dikelurakan untuk melakukan promosi iklan dengan harga Rp 317.050,37. Dari

penelitian diatas bisa dibuktikan bahwa metode estimasi sangat berperan penting pada industri. Dengan melakukan estimasi bisa memprediksi biaya dan keperluan penyelenggaraan training[14].

### 3.3 Estimasi Durasi Promosi Iklan Berbayar

Melakukan analisis estimasi durasi promosi iklan berbayar untuk menentukan durasi perhari promosi iklan. Untuk itu diperlukan modeling menggunakan algoritma estimasi. Dengan membandingkan beberapa algoritma estimasi seperti Network, Linear regression, Support Vector Machine, Generalized Model dan Deep Learning dan melakukan pengujian untuk menghasilkan algoritma terbaik dalam melakukan modelling estimasi durasi iklan berbayar. Setelah melakukan perbandingan algoritma ditentukan algoritma terbaik berdasarkan hasil nilai root mean square error (RMSE) yang menghasilkan nilai mendekati 0 merupakan algoritma terbaik, sehingga dihasilkan algoritma terbaik setelah melakukan uji validasi adalah Linear Regression dengan nilai evaluasi yang dihasilkan sebesar 1.690. Berikut hasil perbandingan algoritma yang sudah didapatkan.

**Tabel 3.** Perbandingan Algoritma untuk melakukan estimasi durasi iklan

Nama Algoritma	Nilai RMSE
Neural Network	2.547
Linear Regression	1.690
Support Vector Machine	2.477
Generalized Linear Model	2.774
Deep Learning	3.115

Setelah menghasilkan algoritma terbaik untuk melakukan pemodelan maka dilakukan modeling menggunakan algoritma linear regression. Karena model linear regression mendekati nilai 0 dan dianggap pendekatan normal[15]. Menggunakan data durasi iklan sebagai parameter sehingga dihasilkan estimasi untuk melakukan promosi iklan yang efisien ialah 6 hari.

## 4. KESIMPULAN

Perbandingan algoritma Generalized Linear Model (GLM), Deep Learning, Neural Network, Linear Regression, dan Support Vector Machine (SVM) untuk estimasi jumlah peserta yang ideal dalam menyelenggarakan suatu training menggunakan algoritma dengan performance terbaik yaitu algoritma Generalized Linear Model (GLM) dengan hasil nilai RMSE (Root Mean Square Error) sebesar 49.984, Support Vector Machine (59.624) menghasilkan nilai 59.624, Deep Learning 71.108, Neural Network dengan nilai 79.243 dan Linear Regression 186.725. Dan menghasilkan jumlah peserta yang efektif dan ideal adalah sekitar 127 orang untuk menyelenggarakan training. Estimasi jumlah biaya iklan dengan menganalisis perbandingan teknik estimasi yaitu Generalized Linear Model (GLM), Deep Learning, Neural Network, Linear Regression, dan Support Vector Machine (SVM) dan algoritma yang memiliki akurasi terbaik menggunakan acuan nilai RMSE (Root Mean Square Error) adalah Neural Network dengan Hasil nilai 112929.701. Sedangkan estimasi yang didapatkan untuk jumlah biaya iklan ialah sekitar Rp 317.050,37. Prediksi lama durasi iklan dalam melakukan promosi berbayar dengan membandingkan algoritma Generalized Linear Model (GLM), Neural Network, Linear Regression, Support Vector Machine (SVM) dan Deep Learning, untuk dijadikan acuan proses estimasi. Hasil algoritma terbaik dalam analisis perbandingan algoritma adalah algoritma Linear Regression dengan hasil nilai evaluasi pada RMSE (Root Mean Square Error) sebesar 1.690. Sedangkan hasil yang didapatkan bahwa lama durasi iklan (dalam hari) yang efektif untuk melakukan promosi adalah sekitar 6 hari. Dari penelitian ini didapatkan hasil bahwa algoritma yang paling baik untuk melakukan estimasi pada suatu data dan untuk estimasi pada suatu kasus tertentu yaitu adalah algoritma neural network dengan hasil uji validasi sebesar 112929.701, kemudian algoritma linear regression dengan nilai validasi sebesar 1.690 dan juga generalized linear model (GLM) juga termasuk algoritma terbaik dengan hasil nilai validasi akurasi sebesar 49.984.

## REFERENCES

- [1] N. Nahar dan F. Ara, "Liver Disease Prediction by Using Different Decision Tree Techniques," *International Journal of Data Mining & Knowledge Management Process*, vol. 8, no. 2, hlm. 01-09, Mar 2018, doi: 10.5121/ijdkp.2018.8201.
- [2] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Rekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, dan S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi".
- [3] P. Handayani *dkk.*, "PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN METODE DECISION TREE DAN NEURAL NETWORK," 2019.
- [4] W. Willy, D. P. Rini, dan S. Samsuryadi, "Perbandingan Algoritma Random Forest Classifier, Support Vector Machine dan Logistic Regression Classifier Pada Masalah High Dimension (Studi Kasus: Klasifikasi Fake News)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 4, hlm. 1720, Okt 2021, doi: 10.30865/mib.v5i4.3177.
- [5] "Klasifikasi pada Dataset Penyakit Hati Menggunakan Algoritma Support Vector Machine, K-NN, dan Naïve Bayes".

- [6] H. Azis, P. Purnawansyah, F. Fattah, dan I. P. Putri, “Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung,” *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, hlm. 81–86, Agu 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [7] R. Nofitri dan N. Irawati, “ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER,” *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, hlm. 199–204, Jul 2019, doi: 10.33330/jurteks.v5i2.365.
- [8] F. S. Nugraha, M. J. Shidiq, dan S. Rahayu, “ANALISIS ALGORITMA KLASIFIKASI NEURAL NETWORK UNTUK DIAGNOSIS PENYAKIT KANKER PAYUDARA,” *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, hlm. 149–156, Agu 2019, doi: 10.33480/pilar.v15i2.601.
- [9] J. P. Matematika, D. Matematika, T. N. Padilah, dan R. I. Adam, “ANALISIS REGRESI LINIER BERGANDA DALAM ESTIMASI PRODUKTIVITAS TANAMAN PADI DI KABUPATEN KARAWANG”.
- [10] F. H. Hamdanah dan D. Fitriana, “Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 1, hlm. 23, Apr 2021, doi: 10.23887/janapati.v10i1.31035.
- [11] B. Pradito dan D. Silvi Purnia, “KOMPARASI ALGORITMA LINEAR REGRESSION DAN NEURAL NETWORK UNTUK MEMREDIKSI NILAI KURS MATA UANG,” *Jurnal Sains dan Manajemen*, vol. 10, no. 2, 2022.
- [12] F. N. Dhewayani *dkk.*, “Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *Jurnal Teknologi dan Informasi*, doi: 10.34010/jati.v12i1.
- [13] Y. Suhandi, I. Kurniati, dan S. Norma, “Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik,” *Jurnal Teknologi Informatika dan Komputer*, vol. 6, no. 2, hlm. 12–20, Sep 2020, doi: 10.37012/jtik.v6i2.299.
- [14] A. Devi Putri Ariyanto dan U. Laili Yuhana, “ANALISIS METODE ESTIMASI BIAYA PADA PERANGKAT LUNAK BESERTA FAKTOR-FAKTOR YANG MEMPENGARUHI: A SYSTEMATIC LITERATURE REVIEW,” vol. 9, no. 4, hlm. 699–708, 2022, doi: 10.25126/jtiik.202294611.
- [15] N. G. Ramadhan, Y. S. R. Nur, dan F. D. Adhinata, “Pendekatan Deep Learning Untuk Prediksi Durasi Perjalanan,” *Teknika*, vol. 11, no. 2, hlm. 85–89, Jun 2022, doi: 10.34148/teknika.v11i2.460.