

Pemodelan Prediksi Kuantitas Penjualan Mainan Menggunakan *Light Gradient Boosting Machine*

Erfan Febriantoro^{a*}, Endang Setyati^b, Joan Santoso^c

^{abc}Pascasarjana Teknologi Informasi Institut Sains Dan Teknologi Terpadu Surabaya, Indonesia
correspondence authors : *erfan.febriant@gmail.com

Abstract— The main characteristic of the toy industry is its fast-changing and unpredictable nature. Demand, influenced by certain trends, can suddenly shift and disappear as the next viral product takes over the market. Short product life cycles, constant product innovation, and high levels of competitive demand can potentially result in relatively higher costs compared to other industries in terms of obsolete inventory, lost sales, and price declines. Based on these issues, a research was conducted to predict toy sales using the LightGBM algorithm model in a time-series format with a sales dataset of 460 toy items classified into 14 categories over a period of 1,353 days with a prediction period of 1, 3, and 6 months. The results of this study presented 42 models based on product categories and prediction periods, with the best RMSE value of 0.0042 for the KARTU toy model, and 3 models for all categories based on prediction periods, with the best RMSE value of 0.0380 for a 1-month prediction period.

Index Terms— LightGBM; Prediction; Sales; Toys.

Abstrak— Ciri khas utama industri mainan adalah perubahan dan ketidakpastiannya yang cepat. Permintaan yang dipengaruhi oleh tren tertentu, dapat berubah secara mendadak dan tiba-tiba menghilang saat produk viral berikutnya menguasai pasar. Siklus hidup yang pendek, inovasi produk yang konstan, dan tingkat kanibalisme yang tinggi berpotensi mengeluarkan biaya relatif lebih tinggi dibandingkan dengan industri lain dalam hal persediaan usang, kehilangan penjualan, dan penurunan harga. Berdasarkan permasalahan tersebut maka dilakukan penelitian untuk memprediksi penjualan mainan menggunakan model algoritma LightGBM dalam bentuk *time-series* dengan dataset penjualan 460 item mainan yang diklasifikasikan menjadi 14 kategori dalam rentang waktu 1.353 hari dengan periode prediksi 1, 3, dan 6 bulan. Penelitian ini menghasilkan 42 model berdasarkan kategori produk dan periode prediksi, dengan nilai RMSE terbaik 0.0042 pada model mainan KARTU, dan 3 model untuk semua kategori berdasarkan periode prediksi dengan nilai RMSE terbaik 0.0380 pada periode prediksi 1 bulan.

Kata Kunci— Mainan; LightGBM; Penjualan; Prediksi.

I. PENDAHULUAN

Mainan adalah salah satu produk konsumsi tertua di dunia. Selama empat dekade terakhir, industri mainan terus mengalami perkembangan dari industri rumah tangga menjadi sebuah industri besar dengan omzet penjualan mencapai USD 5 miliar [1]. Ciri khas utama bisnis mainan adalah perubahan dan ketidakpastiannya yang cepat. Permintaan yang dipengaruhi oleh tren tertentu, dapat berubah mendadak dalam semalam dan kemudian tiba-tiba menguap saat produk viral berikutnya menyapu pasar. Inovasi produk yang konstan, siklus hidup yang pendek, dan tingkat kanibalisme yang tinggi adalah ciri khas industri mainan [1].

Industri mainan berpotensi mengeluarkan biaya relatif lebih tinggi dibandingkan dengan industri lain dalam hal persediaan usang, kehilangan penjualan dan penurunan harga. Hal ini merupakan konsekuensi unik pada *supply chain* mainan, mirip dengan industri pakaian [2]. Untuk bertahan, industri ini menghadapi tantangan yang sangat unik, yaitu menyediakan mainan yang tepat dengan jumlah yang tepat di toko yang tepat selama periode tertentu dan agar dapat menyediakan mainan yang kreatif namun dengan harga yang kompetitif.

Saat ini, arus informasi sama pentingnya dengan arus barang, sehingga perusahaan perlu beradaptasi dengan kecepatan dan keakuratan data yang sangat penting untuk mencapai tujuan bisnis [3]. Transformasi Digital telah menjadi perhatian khusus dan masalah strategis pada semua jenis organisasi [4]. Teknologi digital menawarkan solusi berbeda yang memenuhi tantangan *supply chain* perusahaan dan ekspektasi pasar [5]. Faktanya, pengambilan keputusan *supply chain* umumnya didasarkan pada prediksi penjualan [3].

Machine learning menawarkan kemampuan untuk membuat model dari volume besar data internal dan eksternal. Setelah model-model ini dilatih dan diuji, sistem akan dapat mengantisipasi evolusi sistem, dan akan mengkonfigurasi ulang diri mereka sendiri sesuai dengan model yang terbentuk tersebut. Selain itu, pengambil keputusan melalui machine learning dapat mengidentifikasi variabel baru dan relevan untuk proses operasional mereka [3].

LightGBM adalah framework peningkatan gradien berdasarkan pohon keputusan untuk meningkatkan efisiensi model dan mengurangi penggunaan memori. Dengan mengimplementasikan dua teknik baru yaitu Gradient-based One Side Sampling (GOSS) dan Exclusive Feature Bundling (EFB) yang memenuhi batasan algoritma berbasis histogram yang terutama digunakan di semua kerangka kerja GBDT (Gradient Boosting Decision Tree) [6][14]. Sebagai salah satu dari tiga implementasi GBDT yang populer, LightGBM tidak hanya memiliki keunggulan training speed yang lebih baik, tetapi juga memiliki akurasi prediksi yang tinggi dan lebih hemat memori [7][14][15].

Pada penelitian terdahulu [8], [9], LightGBM terbukti lebih baik dalam penanganan data *time-series* sehingga sesuai dengan ciri khas utama bisnis mainan dengan perubahan dan ketidakpastian yang cepat.

Penelitian ini diharapkan dapat menemukan model yang tepat digunakan untuk memprediksi jumlah penjualan mainan dengan menggunakan metode *machine learning* dalam menganalisis jumlah penjualan mainan dalam bentuk *time series*. Penelitian ini akan membuat model setiap kategori mainan yang terjual menggunakan metode LightGBM dengan periode prediksi dalam 1 bulan, 3 bulan, dan 6 bulan ke depan. Peneliti tertarik menggunakan metode ini karena terbukti lebih baik dalam membuat model prediksi penjualan pada penelitian sebelumnya [8].

II. TINJAUAN PUSTAKA

Pada tahun 2021, [8] melakukan penelitian untuk membuat peramalan penjualan Wal-Mart menggunakan kombinasi metode LSTM dan LightGBM. Dataset yang digunakan adalah data transaksi penjualan dalam bentuk *time-series* selama 1913 hari untuk memprediksi penjualan 28 hari ke depan. Hasil perhitungan RMSSE, menunjukkan bahwa model berbasis LSTM dan LightGBM memiliki kemampuan prediksi yang lebih baik (0,812) dibanding Linear Regression (1,46) dan SVM (1,12).

Di periode yang sama, [9] melakukan penelitian menggunakan dataset yang sama dengan [8] namun hanya menggunakan metode LightGBM. Hasilnya menunjukkan bahwa akurasi LightGBM dengan perhitungan RMSE lebih baik (0,641) dibanding dengan Logistic Regression (0,803) dan SVM (0,732).

Pada tahun 2020, [10] melakukan penelitian pada perusahaan farmasi untuk mendapatkan model peramalan penjualan terbaik dengan membandingkan antara model yang dihasilkan dari kombinasi metode ARIMA-LSTM dengan model tunggal ARIMA dan LSTM. Dataset berasal dari kumpulan data kontes Kaggle yang menyajikan data penjualan harian dari tanggal 2 Januari 2014 hingga 8 Oktober 2019, yang diperoleh dari sistem penjualan farmasi, dengan mengelompokkan data *time-series* menjadi 3 (tiga) bagian berdasarkan sifat penjualannya yaitu N02BA (bersifat menurun), N02BE (musiman/ tahunan), dan N05B (fluktuatif). RMSE dan MAPE dipilih untuk mengevaluasi kinerja prediksi model selama 20 minggu kedepan, keduanya menunjukkan bahwa prediksi menggunakan model kombinatorial pada dasarnya lebih baik daripada model tunggal, baik untuk data temporal dengan keacakan yang lebih besar maupun untuk data temporal yang lebih stabil.

Pada tahun 2020, [11] melakukan penelitian pada perusahaan e-commerce skala ritel online dengan membandingkan antara model yang dihasilkan dari kombinasi metode AGA-LSTM dengan model tunggal LSTM tradisional dan Deep BP. Dataset ritel online yang dipilih sebagai data penelitian berasal dari database UCI, yang mencatat semua informasi pesanan transaksi online ritel online di Inggris dari 1 Desember 2010 hingga 9 Desember 2011. Hasil nilai MSE menunjukkan bahwa nilai error prediksi pada model AGA-LSTM lebih kecil dibandingkan dengan model BP dan LSTM tradisional. Namun, dengan bertambahnya tahapan (step) prediksi, errornya makin besar atau akurasinya makin menurun. Oleh karena itu, peneliti menyarankan untuk lebih fokus pada penelitian prediksi multi-step pada penelitian berikutnya.

Pada tahun 2019, [12] melakukan penelitian untuk menemukan model yang tepat digunakan untuk memprediksi jumlah penjualan obat dengan membandingkan metode kombinasi CNN-LSTM dengan metode tunggal MLP dan LSTM dengan menganalisis jumlah penjualan obat dalam bentuk *time-series*, periode 1 bulan, 3 bulan dan 6 bulan. Dataset berasal dari RS Singburi tercatat sebanyak 1.651 data obat dan terdapat 72 record setiap obat mulai Oktober 2009 sampai September 2015. Obat-obatan diklasifikasikan menggunakan analisis ABC/VEN. "ABC" adalah klasifikasi obat berdasarkan jumlah penggunaan obat dan "VEN" mengklasifikasikan obat menurut kepentingannya dalam pengobatan. Peneliti menggunakan teknik Rolling Window dalam mengekstrak fitur dari data *time-series*, dengan mengulangi regresi menggunakan subsampel dari total data dengan menggeser titik awal dan akhir dengan jendela tetap.

Hasil penelitian menunjukkan bahwa CNN-LSTM menghasilkan akurasi yang lebih baik, hasil juga menunjukkan bahwa periode peramalan 1 bulan cocok untuk obat-obatan yang spesifik untuk penyakit. Periode peramalan 3 bulan cocok untuk obat-obatan yang umum digunakan. Periode ramalan 6 bulan cocok untuk obat-obatan untuk penyakit kronis.

III. METODE PENELITIAN

LightGBM adalah algoritma *Machine Learning* yang dikembangkan oleh Microsoft, yang digunakan untuk masalah pembelajaran mesin berbasis klasifikasi dan regresi. LightGBM mengoptimalkan algoritma Gradient Boosting Decision Tree (GBDT) dengan menggunakan teknik pemotongan yang efisien dan pengelompokan fitur untuk meningkatkan performa dan mengurangi waktu pelatihan. Dengan mengimplementasikan dua teknik baru yaitu Gradient-based One Side Sampling (GOSS) dan Exclusive Feature Bundling (EFB) yang memenuhi batasan algoritma berbasis histogram yang terutama digunakan di semua kerangka kerja GBDT [6]. Sebagai salah satu dari tiga implementasi GBDT yang populer, LightGBM tidak hanya memiliki keunggulan training speed yang lebih baik, tetapi juga memiliki akurasi prediksi yang tinggi dan lebih hemat memori [7].

LightGBM menggunakan teknik pembelahan yang dinamakan leaf-wise growth strategy [9]. Teknik inilah yang akan membatasi kedalaman dari suatu model LightGBM, dan mencari leaf-node yang memiliki splitting gains terbesar, memecah node tersebut, lalu meneruskan proses untuk node yang baru. Level tambahan tidak lagi diperlukan untuk meningkatkan *purity level* atau tingkat kemurniaan dari model, sehingga mencegah model LightGBM tumbuh terlalu dalam. Sebagai hasilnya, sistem akan jauh lebih sedikit mengkonsumsi daya komputasi, dan juga menghindari overfitting [13].

Gradient-based One-Side Sampling (GOSS) merupakan teknik optimalisasi pada algoritma LightGBM untuk mengurangi biaya komputasi dari Gradient Descent yang digunakan pada Gradient Boosting. Teknik ini dikembangkan untuk mengurangi waktu komputasi dan penggunaan memori pada model yang terlalu besar [6]. Cara kerja GOSS adalah dengan mempertahankan semua data pelatihan yang memiliki gradien kuat (strong gradient) dan hanya mempertahankan sebagian kecil data pelatihan yang memiliki gradien lemah (weak gradient). Data pelatihan dengan gradien lemah akan dihapus sebagian secara acak sehingga jumlah data pelatihan dapat berkurang secara signifikan tanpa mengurangi performa model. Dengan cara ini, GOSS memungkinkan model untuk lebih fokus pada data pelatihan yang penting, mengurangi waktu komputasi, dan menghasilkan model yang lebih cepat dan lebih efisien khususnya pada data dengan jumlah fitur yang besar dan jumlah data yang sangat banyak.

Adapun algoritma GOSS adalah sebagai berikut:

```
Input: Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 
      Learning rate  $\alpha$ , number of leaves  $L$ , maximum depth of trees  $max\_depth$ , and other
      parameters
      Subsample ratio of the training instances  $subsample\_ratio$ 
      Ratio of the number of selected instances in subsample  $S$  to the total number of
      instances  $k_s$ 
```

1. Melatih model dengan menggunakan subset sampel D dengan ukuran yang lebih kecil :
 $S = D.sample(subsample_ratio)$
 $F_S = LightGBM(S, \alpha, L, max_depth, \dots)$
2. Menghitung gradien pada seluruh data D :
 $Grad = \nabla_y(F_D)$
3. Mengurutkan gradien mulai dari yang terbesar :
 $Grad_sort = sort(Grad, key=abs(Grad), reverse=True)$
4. Menghitung ukuran subset data yang akan diambil :
 $size_keep = int(k_s * len(D))$
5. Menghitung subset data dengan ukuran sesuai hasil perhitungan sebelumnya, yang terdiri dari instans-instans dengan gradien tertinggi:
 $S_k = Grad_sort[:size_keep]$
6. Sisa data diambil secara acak berdasarkan proporsi gradien setiap instans:
 $S_{ks} = random_sample(Grad_sort[size_keep:], k=len(D)-size_keep, p=|Grad|/\sum|Grad|)$
7. Model akhir dilatih menggunakan subset data yang telah terpilih:
 $F_{S_k} = LightGBM(S_k \cup S_{ks}, \alpha, L, max_depth, \dots)$

Exclusive Feature Bundling (EFB) adalah teknik pengoptimalan pada algoritma LightGBM untuk mengurangi biaya komputasi dan memori pada model yang memiliki banyak fitur. EFB bertujuan untuk menemukan kelompok fitur yang berkorelasi tinggi dan membangun fitur baru yang merupakan gabungan atau bundling dari fitur-fitur tersebut. Cara kerja EFB adalah dengan melakukan eksplorasi untuk menemukan pasangan-pasangan fitur yang memiliki korelasi tinggi menggunakan Correlation-based

Feature Selection (CFS). Setelah ditemukan, pasangan-pasangan fitur tersebut digabungkan menjadi satu fitur baru dengan menggunakan teknik *binning*.

Teknik *binning* dilakukan dengan mengubah setiap nilai pada pasangan-pasangan fitur menjadi label bin. Selanjutnya, fitur baru yang merupakan gabungan dari pasangan-pasangan fitur tersebut dibangun dengan menggunakan pengkodean label bin menjadi variabel baru dalam model. Fitur baru ini selanjutnya digunakan sebagai input pada algoritma LightGBM. Dengan cara ini, EFB dapat mengurangi dimensi data yang besar, menghilangkan fitur-fitur yang tidak penting atau memiliki korelasi yang rendah, serta membangun fitur baru yang lebih berkualitas dan berdampak pada peningkatan performa model. EFB dapat membantu meningkatkan kecepatan dan efisiensi dari algoritma LightGBM, khususnya pada data dengan jumlah fitur yang sangat banyak.

Adapun algoritma EFB adalah sebagai berikut:

Input: **Dataset D dengan n sampel dan p fitur**
Jumlah group g untuk EFB

1. Bagi dataset D menjadi g grup acak yang tidak tumpang tindih, masing-masing berisi :
 $k = p/g$ fitur
2. Untuk setiap grup, hitung average absolute correlation (AAC) antara semua pasang fitur. Correlation matrix harus dihitung terlebih dahulu.
3. Dalam setiap grup, gabungkan pasangan fitur yang memiliki AAC tertinggi hingga jumlah fitur dalam grup menjadi $k/2$.
4. Gabungkan semua pasangan fitur dari setiap grup yang memiliki AAC tertinggi hingga jumlah fitur menjadi k.
5. Buat dataset baru dengan fitur-fitur gabungan tersebut dan lakukan pemodelan dengan algoritma Machine Learning.

Output: **Dataset baru yang berisi fitur-fitur gabungan hasil dari EFB**

Dataset yang digunakan pada penelitian ini berasal dari data penjualan produk mainan CV. Kawi yang beralamat di Kabupaten Jember, yang merupakan perusahaan distributor makanan ringan dan mainan dengan wilayah kerja di kabupaten Jember, Bondowoso, Situbondo, Banyuwangi, Lumajang, Probolinggo, Pasuruan dan Malang. Dataset penjualan mainan yang dipilih sebagai data penelitian adalah data transaksi selama 1.353 hari, sebanyak 115.244 baris data transaksi yaitu mulai tanggal 20 Mei 2018 hingga 31 Januari 2022. Data harian ini diberikan label dengan D_0001 sampai dengan D_1353. Adapun skenario pengujian penggunaan data training dan data tes didasarkan pada periode harian transaksi penjualan adalah sebagai berikut:

Tabel 1 Periode Data Training & Data Tes Penelitian

Periode prediksi	Data Training	Data Test
1 Bulan (30 hari)	D_0001 s.d D_1323	D_1324 s.d D_1353
3 Bulan (90 hari)	D_0001 s.d D_1263	D_1264 s.d D_1353
6 Bulan (180 hari)	D_0001 s.d D_1173	D_1174 s.d D_1353

Dari segi jumlah item barang, terdapat 460 item yang diklasifikasikan menjadi 14 kategori produk berdasarkan pengelompokan yang digunakan pada database perusahaan, dengan kuantiti transaksi penjualan dikonversikan menjadi satuan terkecil.

Pada penelitian ini, setiap kategori tersebut akan dibuatkan 3 (tiga) model berdasarkan periode prediksi pada tabel 1, sehingga secara keseluruhan akan terbentuk 42 model. Secara global juga akan dibuatkan 3 (tiga) model untuk semua kategori berdasarkan periode prediksinya.

Dari uraian data di atas, maka setiap kategori mainan memiliki 1.353 deret data *time-series*, adapun struktur dataset tersebut dapat diilustrasikan pada tabel berikut:

Tabel 2 Struktur Dataset Penjualan Mainan

No	Kategori	Atribut <i>Time Series</i>					
		D_0001	D_0002	D_0003	D_0004	...	D_1353
1	KARTU	0	0	0	12	...	12
2	BUKU & KOMIK	0	0	0	0	...	0
3	BONEKA & ANIMAL	0	0	0	0	...	200
4	PUZZLE	0	360	72	792	...	0
5	MAINAN TRADISIONAL	1	1720	1420	1720	...	600
...
14	SENJATA MAINAN	144	320	0	1160	...	1472

Setiap atribut *time series* memiliki 14 (empat belas) fitur yang digunakan sebagai variabel input, yaitu:

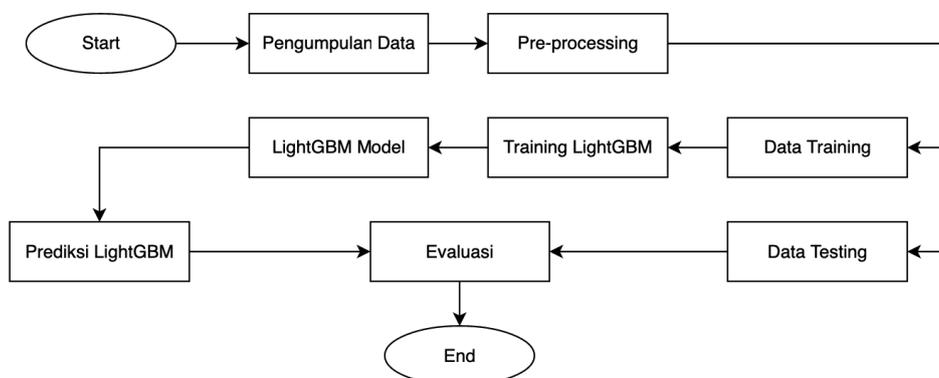
- 1) *One Third Month* (OT), terdiri dari 3 (tiga) kategori yaitu bernilai 1 untuk 10 hari pertama, 2 untuk 10 hari kedua, dan 3 untuk hari ketiga dalam suatu bulan;
- 2) Pekan Rute (PR), terdiri dari 4 (empat) kategori yaitu bernilai 1 untuk pekan pertama, 2 untuk pekan kedua, 3 untuk pekan ketiga, dan 4 untuk pekan keempat/ kelima dalam suatu bulan;
- 3) Libur Sekolah SD (LD) sesuai kalender pendidikan, yaitu bernilai 1 jika ya, 0 jika bukan;
- 4) Libur Sekolah SMP (LP) sesuai kalender pendidikan, yaitu bernilai 1 jika ya, 0 jika bukan;
- 5) Libur Sekolah SMA (LA) sesuai kalender pendidikan, yaitu bernilai 1 jika ya, 0 jika bukan;
- 6) Ujian Sekolah SD (UD) sesuai kalender pendidikan, yaitu bernilai 1 jika ya, 0 jika bukan;
- 7) Ujian Sekolah SMP (UP) sesuai kalender pendidikan, yaitu bernilai 1 jika ya, 0 jika bukan;
- 8) Ujian Sekolah SMA (UA) sesuai kalender pendidikan, yaitu bernilai 1 jika ya, 0 jika bukan;
- 9) Libur Hari Besar Keagamaan (HA) sesuai kalender pemerintah, yaitu bernilai 1 jika ya, 0 jika bukan;
- 10) Libur Hari Besar Nasional (HN) sesuai kalender pemerintah, yaitu bernilai 1 jika ya, 0 jika bukan;
- 11) Puasa Ramadhan (PU) sesuai kalender pemerintah, yaitu bernilai 1 jika ya, 0 jika bukan;
- 12) Libur Idul Fitri/ lebaran (LF) sesuai kalender pemerintah, yaitu bernilai 1 jika ya, 0 jika bukan;
- 13) Libur Natal & Tahun Baru (LT) sesuai kalender pemerintah, yaitu bernilai 1 jika ya, 0 jika bukan;
- 14) PPKM/ PSBB (PM) sesuai kalender pemerintah, yaitu bernilai 1 jika ya, 0 jika bukan.

Adapun fitur *time-series* yang menjelaskan karakter masing-masing atribut atau label mingguan, dijabarkan seperti pada tabel berikut:

Tabel 3 Fitur Time Series

Time Series	OT	PR	LD	LP	LA	UD	UP	UA	HA	HN	PU	LF	LT	PM
D_0001	2	4	1	1	1	1	1	1	0	0	1	0	0	0
D_0002	3	4	0	0	0	1	1	1	0	0	1	0	0	0
D_0003	3	4	0	0	0	1	1	1	0	0	1	0	0	0
D_0004	3	4	0	0	0	1	1	1	0	0	1	0	0	0
...
D_1353	3	4	0	0	0	1	1	1	0	0	1	0	0	0

Berikut adalah diagram dari arsitektur sistem mengenai penelitian yang akan dilakukan :



Gambar 1. Alur Penelitian

Sebelum training dataset dilakukan, terlebih dahulu harus dilakukan proses *preprocessing* data, agar model yang dihasilkan lebih baik dan akurat. Pada penelitian ini teknik *preprocessing* yang digunakan adalah data transformation, yaitu melakukan normalisasi data kuantiti dengan mengubah data menjadi skala yang sama.

Dataset yang telah dilakukan *preprocessing* akan digunakan sebagai data input sesuai dengan skenario pada tabel 1, yaitu dibedakan menjadi *data training* dan *data test* sesuai periode prediksi yang diharapkan. Data training inilah yang digunakan untuk melatih model pada algoritma LightGBM, agar menghasilkan model dengan performa atau akurasi yang baik pada data training itu sendiri dan pada data yang belum pernah dilihat sebelumnya (data test).

Evaluasi akurasi model dihitung menggunakan rumus RMSE (Root Mean Squared Error) dan R-squared (R2). RMSE berfungsi untuk menghitung tingkat kesalahan dalam memprediksi nilai sebenarnya (*actual value*), semakin kecil nilai RMSE, semakin dekat prediksi dengan nilai sebenarnya. Sedangkan R2 berfungsi untuk menjelaskan seberapa baik model dapat menjelaskan variasi (*variance*) dalam data, R2

menghasilkan nilai antara 0 dan 1, dengan nilai 1 menunjukkan bahwa model sepenuhnya menjelaskan variasi dalam data, sementara nilai 0 menunjukkan bahwa model tidak menjelaskan variasi sama sekali.

IV. HASIL DAN PEMBAHASAN

Pengujian model sebagaimana skenario pada tabel 1, menghasilkan model yang dapat dilihat pada tabel 4, dengan akurasi terbaik yaitu pada kategori “**BUKU & KOMIK**” dengan nilai RMSE **0,0048** pada periode prediksi 1 bulan, yang juga terbaik pada periode prediksi 6 bulan dengan nilai RMSE **0,0096**. Sedangkan pada periode prediksi 3 bulan, kategori “**KARTU**” menjadi model terbaik dengan nilai RMSE **0,0094**.

Tabel 4 Hasil Uji Coba Berdasarkan Periode Prediksi

No	Kategori Mainan	Periode Prediksi					
		1 Bulan (30 Hari)		3 Bulan (90 Hari)		6 Bulan (180 Hari)	
		R2	RMSE	R2	RMSE	R2	RMSE
1	KARTU	-0.3933	0.0072	-0.0525	0.0094	-0.1503	0.0132
2	BUKU & KOMIK	-1.1430	0.0048	-4.2576	0.0123	-1.8855	0.0096
3	BONEKA & ANIMAL	0.1939	0.0630	-0.1407	0.0799	0.0145	0.1271
4	PUZZLE	-0.8042	0.0228	-0.2194	0.0268	0.0096	0.0694
5	MAINAN TRADISIONAL	0.0748	0.0809	0.0070	0.0894	0.1125	0.0976
6	MAINAN TERBANG	0.0959	0.1123	0.0622	0.1558	0.0776	0.1308
7	AKSESORIS	0.1330	0.0486	0.1779	0.0568	0.1381	0.0818
8	MAINAN EDUKASI	0.2905	0.1123	0.2121	0.1183	0.1763	0.1016
9	BOLA DAN BALON	0.0255	0.1017	0.1288	0.1653	0.2506	0.1169
10	ALAT SEKOLAH	-0.3937	0.0435	-0.0253	0.0542	0.0648	0.0520
11	ROBOT	-0.7056	0.0385	0.1254	0.0767	0.1621	0.0613
12	ALAT MUSIK	-0.7206	0.0400	0.0773	0.0822	0.0207	0.0793
13	MOBIL & SEPEDA MOTOR	0.2138	0.1655	0.1975	0.1193	0.1260	0.1223
14	SENJATA MAINAN	0.1136	0.1349	0.1209	0.1534	0.0525	0.1327
#	SEMUA KATEGORI	-0.8342	0.0384	0.2304	0.0474	0.3967	0.0660

Hasil akurasi terbaik untuk model “**SEMUA KATEGORI**” adalah pada periode prediksi 1 bulan dengan nilai RMSE **0,0384**, dengan tren akurasi model makin menurun jika rentang waktu periode prediksinya diperpanjang, hal ini ditunjukkan makin besar nilai RMSE pada periode prediksi 3 bulan dan 6 bulan secara berturut-turut.

Hasil perhitungan R-Squared (R2) secara umum menunjukkan bahwa performa model yang dihasilkan kurang baik dalam menjelaskan variasi data (*variance*), artinya model menghasilkan prediksi yang lebih buruk dibanding model *baseline*. Model *baseline* adalah model yang hanya memprediksi nilai rata-rata target untuk semua sampel. Hal ini bisa terjadi jika model yang dibangun terlalu sederhana atau data yang digunakan tidak berkorelasi secara baik.

V. KESIMPULAN

Berdasarkan ujicoba yang telah dilakukan, menghasilkan kesimpulan bahwa metode LightGBM dapat menghasilkan model yang baik untuk memprediksi kuantitas penjualan mainan dalam bentuk data *time-series*, hal ini ditunjukkan dengan nilai RMSE yang baik pada semua periode prediksi, baik secara per kategori produk, maupun secara global. Hasil perhitungan RMSE lebih baik dibanding R-Squared (R2), artinya model cenderung lebih baik dalam memprediksi nilai sebenarnya (*actual value*) daripada dalam menjelaskan variasi data (*variance*).

DAFTAR PUSTAKA

- [1] M. E. Johnson, “Learning from Toys: Lessons in Managing Supply Chain Risk from the Toy Industry,” *California Management Review*, vol. 43, no. 3, pp. 106–124, Apr. 2001, doi: 10.2307/41166091.
- [2] M. Christopher, R. Lawson, and H. Peck, “Creating agile supply chains in the fashion industry,” *Intl J of Retail & Distrib Mgt*, vol. 32, no. 8, pp. 367–376, Aug. 2004, doi: 10.1108/09590550410546188.
- [3] M. Mohamed-Iliasse, B. Loubna, and B. Abdelaziz, “Is Machine Learning Revolutionizing Supply Chain?,” in *2020 5th International Conference on Logistics Operations Management (GOL)*, Rabat, Morocco, Oct. 2020, pp. 1–10. doi: 10.1109/GOL49479.2020.9314713.
- [4] V. Arribas and J. A. Alfaro, “3D technology in fashion: from concept to consumer,” *JFMM*, vol. 22, no. 2, pp. 240–251, May 2018, doi: 10.1108/JFMM-10-2017-0114.
- [5] F. Kache and S. Seuring, “Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management,” *IJOPM*, vol. 37, no. 1, pp. 10–36, Jan. 2017, doi: 10.1108/IJOPM-02-2015-0078.

- [6] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," Dec. 2017. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/>
- [7] F. Ye, J. Wang, Z. Li, Z. Jihan, and C. Yang, "Jane Street Stock prediction model based on LightGBM," in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Xi'an, China, Apr. 2021, pp. 385–388. doi: 10.1109/ICSP51882.2021.9408851.
- [8] H. H. Xie, C. Li, N. Ding, and C. Gong, "Walmart Sale Forecasting Model Based On LSTM And LightGBM," in *2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM)*, Xiamen, China, Jan. 2021, pp. 366–369. doi: 10.1109/ICEKIM52309.2021.00087.
- [9] T. Deng, Y. Zhao, S. Wang, and H. Yu, "Sales Forecasting Based on LightGBM," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, Jan. 2021, pp. 383–386. doi: 10.1109/ICCECE51280.2021.9342445.
- [10] Y. Han, "A forecasting method of pharmaceutical sales based on ARIMA-LSTM model," in *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)*, Shenyang, China, Nov. 2020, pp. 336–339. doi: 10.1109/ISCTT51595.2020.00064.
- [11] K. Chen, "An Online Retail Prediction Model Based on AGA-LSTM Neural Network," in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Taiyuan, China, Oct. 2020, pp. 145–149. doi: 10.1109/MLBDBI51377.2020.00032.
- [12] W. Saena and V. Suttichaya, "Predicting Drug Sale Quantity Using Machine Learning," in *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP)*, Chiang Mai, Thailand, Oct. 2019, pp. 1–6. doi: 10.1109/ISAI-NLP48611.2019.9045222.
- [13] D. Ge, J. Gu, S. Chang, and J. Cai, "Credit Card Fraud Detection Using Lightgbm Model," in *2020 International Conference on E-Commerce and Internet Technology (ECIT)*, Zhangjiajie, China, Apr. 2020, pp. 232–236. doi: 10.1109/ECIT50008.2020.00060.
- [14] D. Wang, L. Li, and D. Zhao, "Corporate finance risk prediction based on LightGBM," in *2022 Information Sciences*, China, Jul. 2022, pp. 259–268. doi: 10.1016/j.ins.2022.04.058.
- [15] L. M. John, R. Shinde, S. Shaikh, and D. Ashar, "Predicting House Prices using Machine Learning and LightGBM." in *2022 SSRN Electronic Journal*, India, Apr. 2022. doi: 10.2139/ssrn.4108744.

Erfan Febriantoro, meraih gelar Sarjana Komputer (S.Kom) dari Universitas Muhammadiyah Jember pada tahun 2016. Saat ini penulis melanjutkan studi Pascasarjana Teknologi Informasi Institut Sains Dan Teknologi Terpadu Surabaya (ISTTS).

Endang Setyati, meraih gelar Insinyur Matematika (Ir.) dari Institut Teknologi Bandung (ITB) pada tahun 1992. Kemudian meraih gelar Master (MT) dari Institut Teknologi Sepuluh Nopember (ITS) di Surabaya pada tahun 1998. Dan meraih gelar Doktor (Dr) dari Institut Teknologi Sepuluh Nopember (ITS) di Surabaya pada tahun 2016. Saat ini penulis menjadi dosen di Institut Sains Dan Teknologi Terpadu Surabaya (ISTTS), bidang peminatan *computer vision*, logika matematika, *digital image processing*.

Joan Santoso, meraih gelar sarjana komputer (S.Kom) dari Sekolah Tinggi Teknik Surabaya (STTS) pada tahun 2011. Kemudian meraih gelar Master (M.Kom) dari Sekolah Tinggi Teknik Surabaya (STTS) pada tahun 2013. Dan meraih gelar Doktor (Dr) dari Institut Teknologi Sepuluh Nopember (ITS) di Surabaya pada tahun 2020. Saat ini penulis menjadi dosen di Institut Sains Dan Teknologi Terpadu Surabaya (ISTTS), bidang peminatan *web mining*, *natural language processing*, *data mining*, *artificial intelligence*, *big data analytics*.