

Prediksi Penjualan Supermarket Menggunakan Pendekatan Deep Learning

Market Sales Prediction Using Deep Learning Approach

Marchel Thimoty Tombeng¹⁾, Zalfie Ardian²⁾

¹Fakultas Ilmu Komputer, Universitas Klabat

²Fakultas Ilmu Komputer, Universitas Ubudiyah Indonesia

e-mail: marcheltombeng@unklab.ac.id¹, Zalfie.ardian@uui.ac.id²

Abstrak

Berdasarkan data transaksi tahun 2014 sampai 2016 dari salah satu supermarket yang ada di Taiwan, penulis menghasilkan analisa model prediksi dengan menguji data menggunakan metode Deep Learning. Beberapa faktor yang berpengaruh telah di dipelajari dan berguna untuk input prediksi, antara lain keadaan cuaca, diskon, hari raya, dan lain sebagainya. Motivasi utama dari penelitian yang penulis lakukan adalah menggunakan teknologi yang berhubungan dengan eksplorasi data untuk memprediksikan penjualan dari produk-produk dan waktu berkunjung pelanggan dalam industry retail, untuk mencari grup target yang tepat dan korelasi produk yang tinggi. Pada akhirnya penulis menciptakan sistem keputusan produk yang berisi analisa visual dan tindakan saran untuk manajer produk pemasaran serta pemangku kepentingan dalam pemasaran produk. Dengan adanya hasil prediksi ini, diharapkan dapat membantu manajer atau pemangku kepentingan lainnya untuk dapat memasarkan serta menjual produk secara tepat sehingga dapat menghasilkan keuntungan yang banyak dengan menggunakan analisa prediksi yang kami buat. LSTM merupakan model yang sering dipakai dalam Recursive Neural Network (RNN), dan pada dasarnya berfungsi untuk memecahkan masalah dari Time Series. Model Deep Learning yang penulis gunakan adalah Long Short Term Memory (LSTM), dimana model ini menyediakan analisa dan prediksi dari serangkaian data. Sebagai contoh, pada saat akhir pekan pengunjungnya melonjat, maka time machine learning ini akan menambahkan pengartian dari nilai parameter akhir pekan dan nilai ouputnya memiliki korelasi yang kuat.

Kata kunci—Predictions, Time Series, LSTM, RNN, Deep Learning

Abstract

Based on transaction data from 2014 to 2016 from one of the supermarkets in Taiwan, the authors produced a predictive model analysis by testing the data using the Deep Learning method. Several influencing factors have been studied and are useful for predictive input, including weather conditions, discounts, holidays, and so on. The main motivation of the research that the author does is to use technology related to data exploration to predict sales of products and customer visiting times in the retail industry, to find the right target group and high product correlation. In the end the authors created a product decision system containing visual analysis and action suggestions for marketing product managers as well as stakeholders in product marketing. With the results of this prediction, it is hoped that it can help managers or other stakeholders to be able to market and sell products appropriately so that they can generate large profits using the predictive analysis we developed. LSTM is a model that is often used in Recursive Neural Networks (RNN), and basically functions to solve problems of the Time Series. The deep learning model that author use is Long Short Term Memory (LSTM), where this model provides

analysis and prediction of a series of data. For example, when the visitor's weekend is increasing, this time machine learning will add the meaning of the weekend parameter value and the output value has a strong correlation.

Keywords— Predictions, Time Series, LSTM, RNN, Deep Learning

1. PENDAHULUAN

Pada masa kini, teknologi pengembangan telah meningkat sangat cepat [1], terutama dalam bidang analisa data. Motivasi utama penelitian ini adalah menggunakan teknologi eksplorasi data untuk memprediksi penjualan barang dan waktu kunjungan pelanggan di industri ritel, menemukan kelompok sasaran yang tepat dan produk yang berkorelasi tinggi untuk mencapai tujuan promosi produk dan mengurangi yang tidak perlu [2].

Berdasarkan data transaksi supermarket tertentu di cabang Taoyuan pada bulan Januari 2014 hingga Oktober 2016. Penelitian ini menganalisis korelasi antara masing-masing produk dengan pelanggan. Mencari tahu perilaku konsumsi mereka untuk menganalisis produk mana yang paling disukai orang, dan selanjutnya digabungkan dengan analisis aturan asosiasi produk, dan melalui hasil analisis. Toko dapat merancang program promosi dan barang-barang promosi yang meningkatkan kemauan pelanggan untuk membeli, dan dapat membantu toko untuk meningkatkan strategi penjualan.

Untuk memahami perilaku, seperti kesukaan, kesetiaan, kinerja penjualan serta lain-lain dari pelanggan terhadap produk, penulis menganalisis catatan transaksi data dari pelanggan. Penulis berharap dapat mempelajari beberapa pengetahuan dari data tersebut dan mencoba menemukan model keuntungan yang layak untuk feedback.

Penulis menggabungkan data produk melalui grup pelanggan. Karena setiap pelanggan memiliki produk kesukaannya, penulis kemudian mengelompokkan pelanggan dengan perilakunya. Penulis menemukan fitur untuk setiap produk dan memeriksa dengan model prediksi yang digunakan. Hasil prediksi membuat manajer mengetahui nilai jual di masa depan. Model deep learning membantu memprediksi hasil penjualan dengan mempertimbangkan beberapa faktor, seperti cuaca atau musim, dan lainnya [3]. Hal ini akan meningkatkan tingkat loyalitas anggota karena mereka mungkin terpengaruh dengan pengaturan yang lebih tepat.

2. METODE PENELITIAN

Dalam beberapa tahun terakhir, karena pengaruh data besar, manajemen data dan teknik analisis menjadi lebih penting [4]. Oleh karena itu, penulis merancang sistem data warehouse, dan membuat analisis pelanggan dan produk menjadi lebih akurat. Untuk mencapai tujuan tersebut, penulis menggunakan beberapa teknik analisis data seperti data mining, sistem pendukung keputusan turunan, dan deep learning untuk prediksi data. Dalam tulisan ini, penulis fokus pada pendekatan deep learning, lebih khusus lagi fokus pada prediksi menggunakan long short term memory. Berikut ini merupakan penjelasan dari penelitian terkait serta metode penelitian yang penulis gunakan.

2.1 Penelitian Terkait

Ada beberapa penelitian terkait yang penulis gunakan untuk memberikan indikator penelitian yang penting untuk membantu penulis memberikan arahan analitis.

Penelitian pertama membahas tentang prediksi tenaga angin berdasarkan Recurrent Neural Network (RNN) dengan unit Long Short-Term Memory (LSTM) [5]. Mereka mengusulkan pemodelan RNN menggunakan LSTM berdasarkan struktur temporal data

kecepatan angin. RNN adalah fitur penting karena dapat meneruskan informasi sebelumnya dari waktu ke waktu, dan untuk jangka waktu yang lama unit LSTM dapat lebih mempertahankan memori. Mereka berfokus pada data angin yang diambil dari Global Energy Forecasting Competition untuk melakukan analisis. Tujuan dari penelitian mereka adalah mengembangkan model untuk meramalkan kecepatan angin dalam berbagai periode waktu berdasarkan data historis masa lalu.

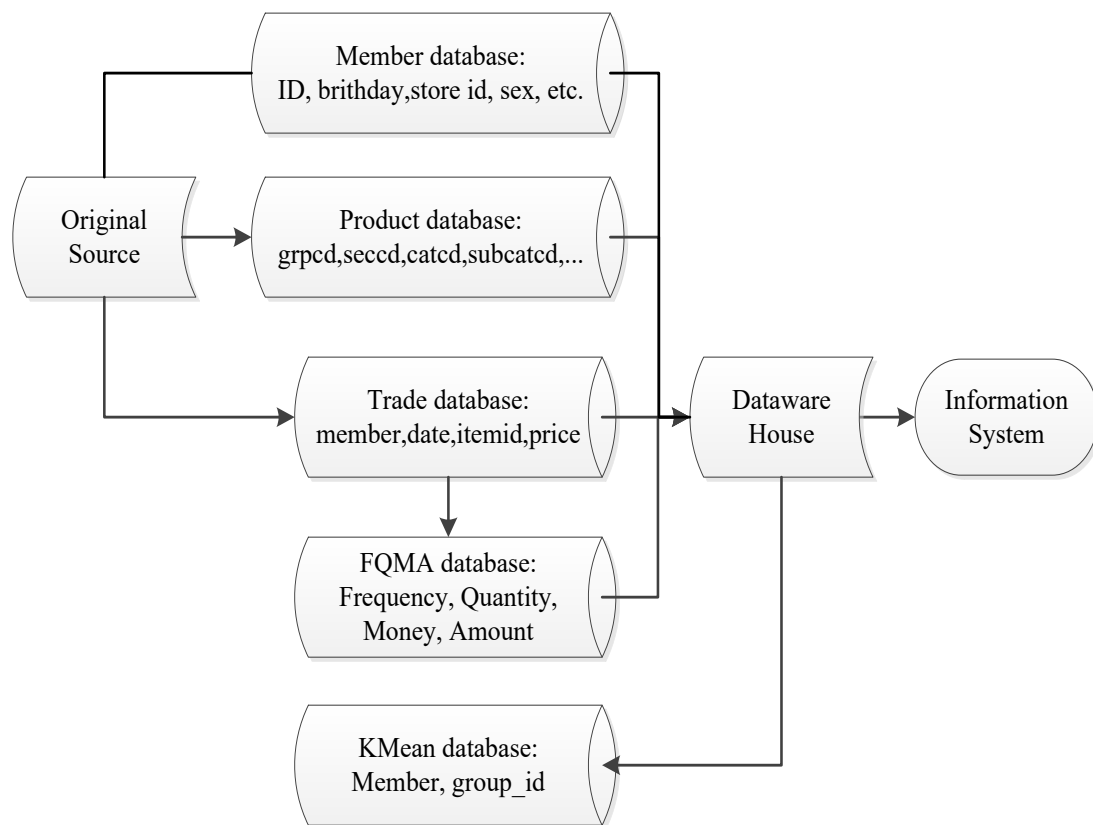
Penelitian kedua membahas prediksi untuk short-term network traffic [6], penulis penelitian ini mencoba menemukan metode mana yang dapat secara efektif mempelajari tren perubahan pada perincian waktu yang sesuai, dan untuk nilai rata-rata lalu lintas di masa mendatang selama periode waktu, prediksi waktu nyata dilakukan. Hasilnya, model prediksi waktu nyata yang dibangun dengan jaringan LSTM diusulkan untuk memprediksi lalu lintas jaringan jangka-bidik. Berdasarkan hasil percobaan perbandingan didapatkan bahwa jaringan LSTM lebih cocok untuk membangun model trafik real-time dibandingkan dengan jaringan syaraf tiruan SVR dan BP, hal ini terjadi karena dapat terus diupdate agar sesuai dengan karakteristik trafik terkini. Selain itu, penelitian ini membahas tentang metode peringatan kemacetan tautan berdasarkan LSTM dan membuktikan bahwa metode tersebut berguna untuk memprediksi terjadinya kemacetan tautan.

Penelitian ketiga membahas tentang teknik long short-term memory yang diimplementasikan pada model prediksi keandalan perangkat lunak dan menambahkan propagasi mundur terpotong dan normalisasi lapisan ke model untuk meningkatkan kinerja [7]. Hasil dari penelitian ini, pendekatan jaringan LSTM yang diusulkan memiliki kinerja prediksi dan ketahanan yang lebih baik dibandingkan dengan model jaringan saraf lainnya.

Penelitian keempat [8] membahas prediksi rating aplikasi app store menggunakan algoritma random forest. Penulis dalam penelitian ini mengimplementasikan teknik data mining dan menggunakan metode Random Forest untuk menganalisa jumlah pengguna yang menyukai sudah dikembangkan yang telah di publish di app store.

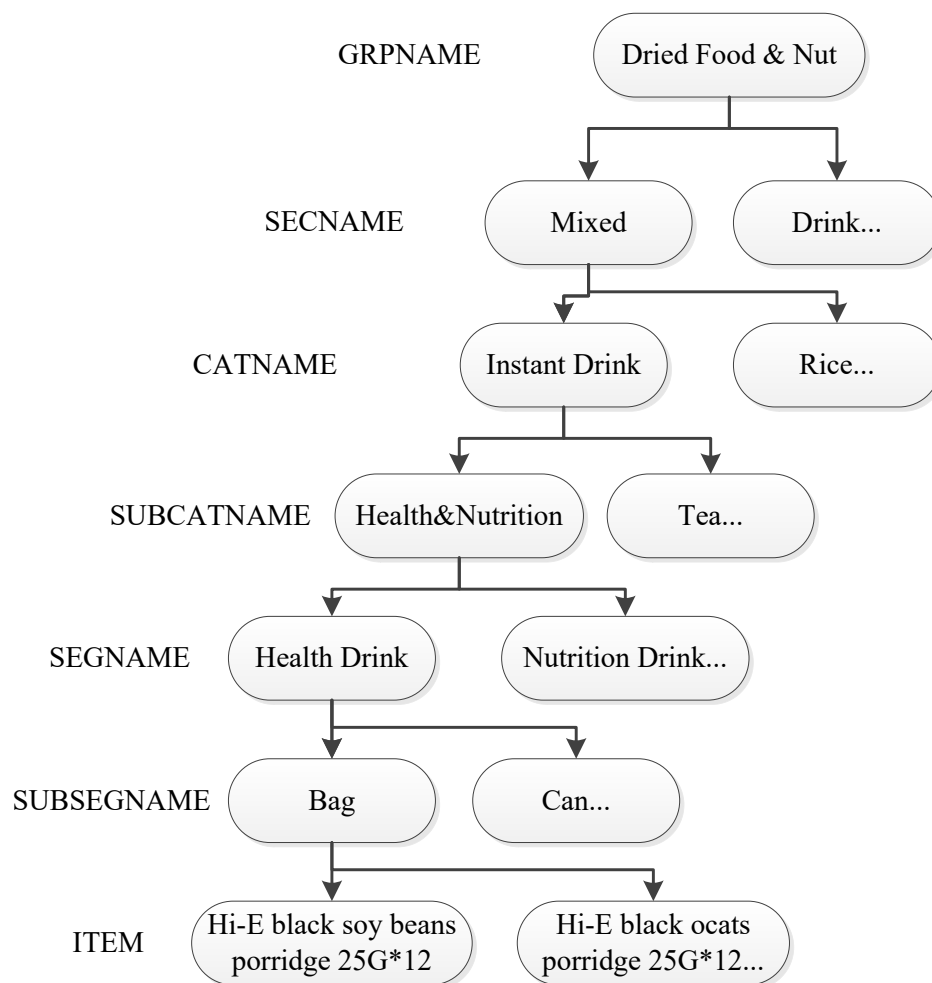
2.2 Analisa Database

Database yang digunakan berasal dari salah satu supermarket di Taiwan. Database pada Gambar 1 terdiri dari data produk, data pelanggan dan data transaksi penjualan. Supermarket memiliki 19 cabang dan jumlah pelanggan yang terdaftar adalah sekitar 6 juta orang. Sebagai contoh, cabang Taoyuan di tahun 2014, terdapat 160,000 member pelanggan, dan rerata member pelanggan yang mengunjungi cabang tersebut adalah 5 sampai 6 kali setiap tahunnya. Terdapat sekitar 630.000 item dalam data produk, dimana hanya sekitar 52.000 item di cabang tersebut yang melakukan promosi penjualan.



Gambar 1 Database Flowchart

Skema Klasifikasi pada Gambar 2 tentang data komoditas dijelaskan sebagai berikut. Ada tiga kategori utama, pertama adalah klasifikasi besar, kedua adalah klasifikasi tipe, dan ketiga adalah klasifikasi penampilan. Setiap klasifikasi dibagi menjadi sub-kategori [9]. Oleh karena itu, bagian klasifikasi memiliki enam lapisan, sedangkan lapisan ketujuh merupakan item produk utama. Karena kuantitas kemasan yang berbeda, produk serupa yang sama akan dibagi menjadi item yang berbeda, tetapi mereka termasuk dalam klasifikasi umum. Lapisan ketujuh adalah penjualan aktual, yang berisi nama produk yang akan dilihat konsumen. Menurut daftar database saat ini, ada lebih dari 630.000 nama item yang berbeda, setiap item memiliki ID independen untuk membedakannya dari item lain.



Gambar 2 Hierarki Komoditas

2.3 Analisa Prediksi LSTM

LSTM adalah model yang paling umum digunakan di Recurrent Neural Network (RNN) [8]. Ini pada dasarnya digunakan untuk memecahkan masalah time series, dan dengan menyimpan keluaran dari lapisan tersembunyi di memori. Ketika data berikut di masukan, nilai dari memori terakhir dipertimbangkan untuk kalkulasi dan LSTM tidak disertakan [10]. LSTM menggunakan Forgetting Gate dan Input Gate unik untuk menambahkan cabang memori selain output yang diprediksi dan memutuskan apakah akan mengingat karakteristik input ini.

Forget Gate diekspresikan dalam ft., jika kata saat ini adalah topik baru atau kata yang berlawanan, maka kata sebelumnya akan dihancurkan oleh Forgetting Gate, dan kebalikan dalam memori. Input Gate menentukan apakah input saat ini (Input) dan unit memori yang baru dibuat (Memory Cell Candidate) ditambahkan ke memori jangka panjang (Long Term Memory).

Penelitian ini menggunakan model long-term dan shor-term memory untuk menyusun model prediksi penjualan komoditas, dan menetapkan indikator prediktif melalui karakteristik barang dan rangkaian waktu penjualan. Sumber data yang penulis gunakan dalam melakukan penelitian ini adalah semua data dari cabang Taoyuan selama tiga tahun terakhir sebagai input. Data yang di gunakan dalam percobaan terdiri dari 66% data pelatihan dan 33% data uji dari informasi penjualan tahun 2014 dan 2015, dan semua data tersebut digunakan sebagai masukan untuk bahan pembelajaran prediksi penjualan tahun 2016.

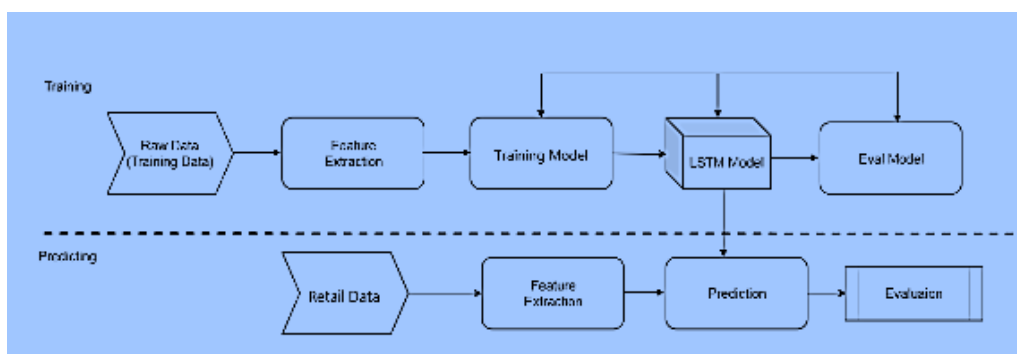
Dalam menganalisis data, kami menerapkan metode pembelajaran mesin untuk memahami waktu dan peluang penting untuk meningkatkan pendapatan. Selain itu, faktor waktu,

pilihan penjualan produk juga menjadi masalah. Penulis tidak dapat secara artifisial dan perlahan menemukan produk unggulan yang cocok dengan musim. Meskipun beberapa produk dapat ditemukan sesuai dengan waktu, penulis membutuhkan rumus dan model yang lebih spesifik dan menggunakan program untuk mendapatkannya jawabannya. Ini lebih cepat, efisien dan dapat mengecualikan spekulasi dan kesalahan perhitungan pekerjaan manual.

Model prediksi penjualan yang dibuat melalui deep learning perlu menunjukkan hari libur penting atau mungkin menunjukkan tanggal kesalahan yang diperbolehkan. Saat mempertimbangkan hari raya, konsumen memiliki kesempatan untuk mengonsumsi produk terkait terlebih dahulu. Keluaran model perlu menunjukkan titik waktu produk yang direkomendasikan dan kenaikannya, sehingga perusahaan dapat memprediksi persediaan waktu dan kuantitas permintaan produk [11]. Dengan model prediksi jumlah konsumsi, komputer dapat mengetahui semua produk penting dari festival yang relevan dan menggunakan produk untuk menghitung rumus, lalu menggunakan rumus tersebut untuk memprediksi generasi dan kesesuaian festival di tahun mendatang. Dan kita dapat menggunakan data historis untuk membandingkan keakuratan prediksi.

Bagian input deep learning harus memperhitungkan faktor-faktor yang dapat mempengaruhi kualitas penjualan. Keuntungan menggunakan deep learning adalah kita dapat mengumpulkan elemen yang memungkinkan untuk membuat penilaian. Menurut analisis eksperimental saat ini, penulis menemukan beberapa kemungkinan pengaruh. Unsur-unsur penjualan meliputi tanggal, akhir pekan, tahun, curah hujan, suhu. Oleh karena itu, setelah mengumpulkan informasi yang relevan, penulis mengubah informasi tersebut menjadi parameter pembelajaran untuk mempelajari parameter mana yang dapat mempengaruhi hasil penjualan, dan terakhir menggunakan pengaruh pembelajaran untuk memprediksi nilai jual dari tahun tersebut.

Dalam deep learning, penulis menggunakan modul LSTM yang disediakan oleh Keras, yang menyediakan model prediktif untuk deep learning, yang menganalisis dan memprediksi serangkaian data. Misalnya, saat akhir pekan sedang besar, maka deep learning akan meningkatkan nilai parameter akhir pekan kognitif dan nilai keluaran memiliki korelasi yang besar. Untuk proses flow dari LSTM dapat dilihat pada Gambar 3.



Gambar 3 Flow Proses LSTM [7]

3. HASIL DAN PEMBAHASAN

Dalam pengolahan data, penulis menormalisasi data yang diprediksi sehingga datanya antara 0 dan 1, dan sekitar 66% data ditetapkan sebagai sisa 33% dari data latih. Data ditetapkan sebagai data pengujian, dan terakhir keluaran model prediksi. Penulis menggunakan perputaran sebelumnya untuk memprediksi periode saat ini, jadi penulis menggunakan keluaran model (pra-konsumsi, konsumsi saat ini) sebagai (X, Y), dan terakhir penulis menggunakan rumus MAPE

yang terlihat pada rumus (1) untuk mengevaluasi efek model setiap komoditas [12]. Konversi data pada data yang di impor ditunjukkan pada Tabel 1 di bawah ini:

Rumus Mean Absolute Percent Error

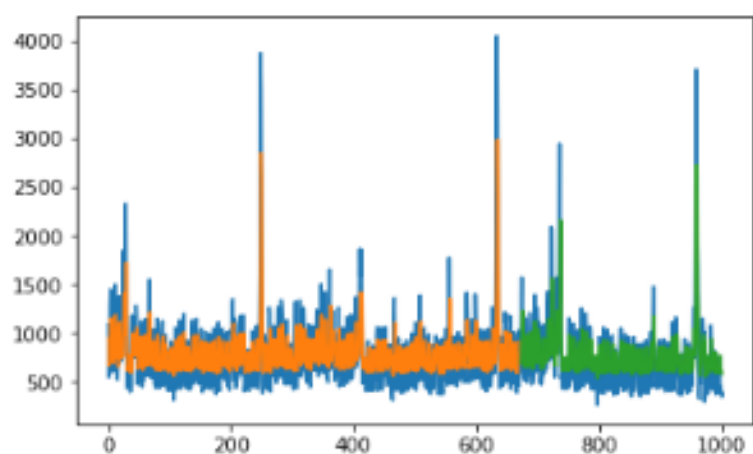
$$MAPE = \frac{\sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i}}{n} \times 100 \quad (1)$$

Tabel 1 Konversi data yang di impor

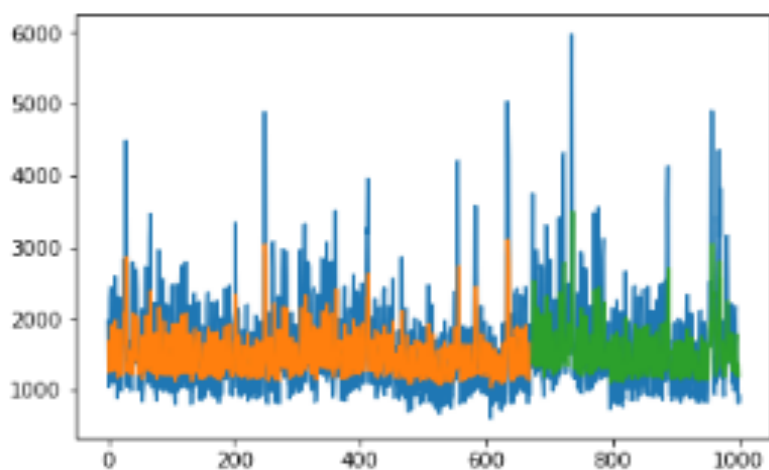
Raw Data(X)	Data Processing(Y)	Note
Month	$y = \text{Modx}(\text{month})$	
Date	$v = \text{Modx}(\text{days})$	
Weekend	Def: Mon, Thu=0; Fri=0.5; Sat Sun=1	
Festival	$y = z(\text{year}) - \text{FestivalDays}$	
Raining	Normalization(y) $= x(\text{daily rainfall})$	If $y \geq 1$, then $v = 1$
Temperature	Normalization(y) $= x(\text{average temperature})$	
Members	Normalization(y) $= x(\text{daily member})$	
Selling	Normalization(y) $= x(\text{daily customer for each product})$	
Revenue	Normalization(y) $= x(\text{daily income for each product})$	

Beberapa hari libur dapat memengaruhi penjualan [13]. Seperti beberapa festival yang sebagian besar didasarkan pada kalender lunar. Oleh karena itu, tanggal kalender nasional akan diubah menurut tahun yang berbeda [14], dan suhu cuaca historis diperoleh melalui situs web Biro Cuaca Pusat. Data historis hanya memiliki suhu bulanan rata-rata dan, bagaimanapun, penulis percaya bahwa ketika curah hujan cukup besar, tidak peduli berapa banyak hujan, ini membuat tidak nyaman bagi orang untuk keluar rumah. Setelah menormalisasi nilai yang dikurangi, penulis membatasi tak terhingga ketika rentang yang sama terlampaui.

Pada Gambar 5 dan 6, penulis memilih sayuran dan produk daging untuk diprediksi dengan banyak penjualan, jumlah pembelian, dan total penjualan. Nilai MAPE antara 5% dan 6%, tetapi jika kita menggunakan produk bir untuk memprediksi nilai MAPE, tidak ada cara untuk menjadi serupa. Ketepatan produk sayuran dan daging berarti bahwa produk bir memiliki keuntungan yang baik di merchandiser massal, tetapi bentuk gelombang penjualannya tidak stabil dengan kebutuhan esensial [15] (sayuran / daging).



Gambar 5 Prediksi Penjualan Class Product Sayuran



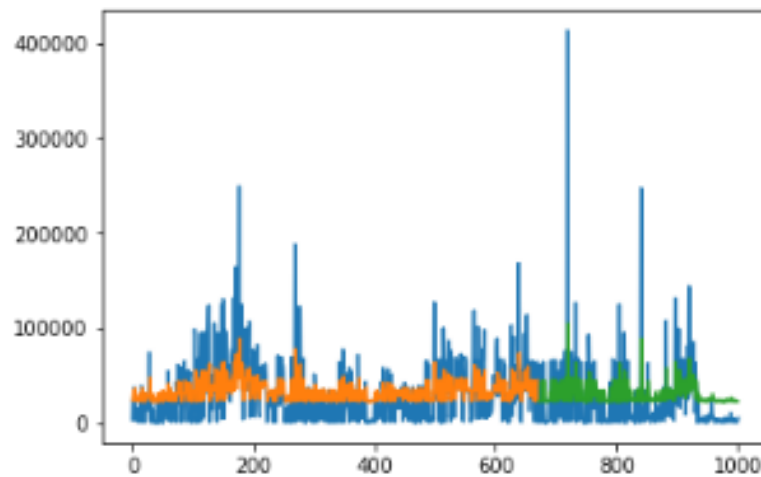
Gambar 6 Prediksi Penjualan Class Product Daging

Hasil prediksi dari nilai-nilai MAPE dijelaskan pada Tabel 2

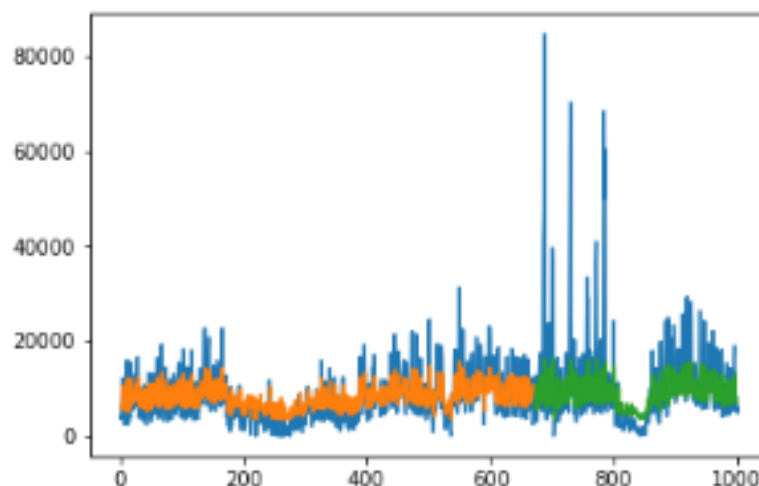
Tabel 2 Hasil prediksi nilai MAPE

Product Model ⁺	Forecasting Type ⁺	MAPE Value ⁺
Vegetable Class Products ⁺	Selling ⁺	5.56 ⁺
Vegetable Class Products ⁺	Purchase ⁺	6.15 ⁺
Vegetable Class Products ⁺	Revenue ⁺	5.39 ⁺
Meat Class Products ⁺	Selling ⁺	5.4 ⁺
Meat Class Products ⁺	Purchase ⁺	5.37 ⁺
Meat Class Products ⁺	Revenue ⁺	5.48 ⁺

Selain itu, penulis menggunakan bir dan susu segar yang terlihat pada Gambar 7 dan 8 sebagai contoh untuk membangun model prediksi komoditas, dengan nilai MAPE 7,2% dan 6,1%.



Gambar 7 Prediksi Penjualan Class Product Bir



Gambar 8 Prediksi Penjualan Class Product Susu Segar

Hasil prediksi LSTM ditunjukkan pada Gambar 5, 6, 7, dan 8 serta pada Tabel 2 di atas, di mana tabel garis biru solid adalah data aktual, garis oranye dan hijau masing-masing direpresentasikan sebagai hasil data pengujian dan prediksi. Meskipun tingkat kesalahan desain model sekitar 5%, penulis tidak secara khusus menguraikan efek musiman melalui dampak prediksi, namun model LSTM ini telah dipertimbangkan untuk penulis setelah 100 putaran pelatihan data.

4. KESIMPULAN

Pada hasil prediksi penjualan, didapati bahwa produk bir memiliki angka prediksi yang baik dibandingkan produk sayuran dan daging, namun gelombang penjualannya relatif tidak stabil dibandingkan dengan kebutuhan esensial seperti produk sayuran dan daging dan menurut analisa penulis ini disebabkan oleh karena produk sayuran dan daging merupakan kebutuhan primer sehingga orang-orang membutuhkannya setiap hari walaupun harganya tidak semahal bir. Namun

untuk produk bir yang mana harganya mahal, sering dibutuhkan oleh orang-orang saat acara-acara tertentu seperti festival ataupun acara-acara tahunan seperti ulang tahun dan lain sebagainya sehingga gelombangnya tidak akan stabil .

Salah satu kendala yang dihadapi yaitu penulis tidak mendapatkan data yang lengkap tentang jumlah promosi iklan, oleh sebab itu jika data-data ini tersedia secara lengkap, maka prediksi serta analisisnya akan lebih akurat jika menggunakan informasi promosi selama tiga tahun terakhir, dan optimasi model prediksi ini bisa lebih efektif dan akurat.

Penulis berharap metode analisis prediksi menggunakan deep learning memungkinkan manajer dan pemangku kepentingan untuk lebih cepat memahami informasi penjualan dan hubungan pelanggan dari toko ritel sehingga rencana menyangkut penjualan bisa di buat agar pendapatan supermarket ini dapat meningkat.

5. SARAN

Untuk penelitian lebih lanjut, saran kami adalah dapat menggunakan data penjualan produk yang terbaru, dan menggunakan variansi *class product* banyak seperti buah-buahan, ikan, dan lain-lainnya agar analisa prediksi bisa lebih lengkap dan akurat, serta menggunakan metode-metode prediksi lainnya untuk melihat perbandingan tingkat akurasi dari tiap-tiap metode prediksi yang ada.

DAFTAR PUSTAKA

- [1] Tombeng, M. T., Kadow, H., Adam, S. I., Silitonga, A., Korompis, A., 2019, Android-Based Application To Detect Drowsiness When Driving Vehicle, *IEEE 1st International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 100-104
- [2] Liang, J. F., 2013, Applying Data Mining Techniques on Retailing Industry – A Case Study on Speciality Chain Store, *Master Thesis*, National Taipei University Department of Business Administration.
- [3] Chang, Y. C., 2015, A Study on Database Marketing Strategy of 3C Chain Stores, *Master Thesis*, National Kaoshiung First University of Science and Technology Department of Marketing and Distribution Management.
- [4] Shih, M. H., 2013, Applying Data Mining to Customer Relationship Management – A study on Construction Materials Retail Business, *Master Thesis*, Tunghai University Department of Industrial Engineering.
- [5] Dong, D., Shen, Z., Yang, T., 2018, Wind Power Prediction based on Recurrent Neural Network with Long Short-term Memory Units, *2018 IEEE International Conference on Renewable Energy and Power Engineering*, pp. 34-38.
- [6] Lu, H., Yang, F., 2018, Research on Network Traffic Prediction Based on Long Short-Term Memory Neural Network, *2018 IEEE 4th International Conference on Computer and Communications*, pp. 1109-1113.

-
- [7] Yangzhen, F., Hong, Z., Chenchen, Z., Chao, F., 2017, A Software Reliability Prediction Model, *2017 IEEE International Conference on Software Quality, Reliability, and Security (Companion Volume)*, pp. 614-615.
- [8] Sandag, G., 2020, Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest, *Cogito Smart Journal Fakultas Ilmu Komputer Universitas Klabat*, vol 6, no 2, pp. 167-178.
- [9] McQueen, J. B., 1967, Some Methods of Classification and Analysis of Multivariate Observations, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistic and Probability*, pp. 281-297.
- [10] LSTM Model, https://en.wikipedia.org/wiki/Long_short-term_memory, diakses tgl 23 Februari 2020.
- [11] Berry, M. J. A, Linoff, G. S., 2011, *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management*, Vol. 1, Ed. 1, Wiley Computer Publishing, Hoboken NJ.
- [12] Charlet, L., Annie, M. C., Kumar, D. A., 2012, Market Basket Analysis for a Supermarket Based on Frequent Item Set Mining, *International Journal of Computer Science Issues*, 9(3), pp. 257.
- [13] Schmitt, J., 2010, Drawing Association Rules between Purchases and In-Store Behavior: An Extension of the Market Basket Analysis, *Advances in Consumer Research*, 37, pp. 899-901.
- [14] Adomavicius, G., Tuzhilin, A., 2005, Incorporating Contextual Information in Recommender System Using a Multidimensional Approach, *ACM Transactions on Information System*, Vol. 23(1), pp. 103-145.
- [15] Kohavi, R., Parekh, R., 2004, Visualizing RFM Segmentation, *Proceedings of the 2004 SIAM International Conference on Data Mining (SDM'2004)*, Orlando, FL, pp. 391-399.