

Prediksi Penjualan Pada Data Penjualan Perusahaan X Dengan Membandingkan Metode GRU, SVR, DAN SARIMAX

Jordan Nagakusuma, Henry Palit, Hans Juwiantho

Program Studi Sistem Informasi Bisnis Fakultas Teknologi Industri Universitas Kristen Petra

Jl. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031) – 2983455, Fax. (031) - 8417658

E-mail: jordan.naga@gmail.com, hnpalit@petra.ac.id, hans.juwiantho@petra.ac.id

ABSTRAK

Sales forecasting merupakan sebuah usaha untuk melakukan prediksi penjualan menggunakan beberapa metode, seperti metode *statistic*, *machine learning*, dan lain – lain. *Sales forecast* dinyatakan penting karena *forecast* yang buruk dapat berdampak pada *inventory* perusahaan sehingga dapat menyebabkan penyimpanan barang terlalu banyak maupun terlalu sedikit sehingga menyebabkan perusahaan rugi. Oleh karena itu, diperlukan sebuah model yang dapat melakukan prediksi penjualan supaya perusahaan dapat melakukan perencanaan sebelum melakukan pengisian stok barang. Namun, *forecast* tidak bisa dilakukan secara langsung, karena data penjualan sebuah perusahaan pasti dipengaruhi oleh berbagai macam faktor dan penjualan bulan yang lalu tidak selalu sama dengan yang akan datang, sehingga perlu adanya data external dalam melakukan prediksi penjualan ke depan.

Oleh karena itu, pada penelitian ini akan dilakukan sebuah prediksi menggunakan 3 model, yaitu model GRU, SVR dan SARIMAX dengan bantuan data external berupa data IHK dan data Inflasi. Selain itu, pada penelitian ini juga dilakukan sebuah uji korelasi untuk mengetahui apakah data penjualan yang akan diprediksi memiliki signifikansi / keterkaitan dengan data external sehingga membantu dalam memprediksi data penjualan.

Hasil yang didapat dari penelitian ini adalah data *pot* lebih cocok menggunakan data univariate dengan model GRU, dengan RMSE Train 3.22, RMSE Test 2.93. Untuk data *hanger* dan *sealware*, model yang paling bagus untuk diterapkan prediksi adalah model SARIMAX dengan jenis data *univariate* (RMSE 30.43) dan dengan jenis data *multivariate* (RMSE 8.07).

Kata Kunci: *Sales Forecasting, GRU, SARIMAX, SVR, IHK, Inflasi, prediksi penjualan*

ABSTRACT

Sales forecasting is an attempt to predict sales using several methods, such as statistical methods, machine learning, and others. Sales forecasts are considered important because poor forecasts can have an impact on the company's inventory so that it can cause storage of too much or too little goods, causing the company to lose. Therefore, we need a model that can predict sales so that companies can plan before filling stock. However, forecasts cannot be done directly, because a company's sales data is definitely influenced by various factors and sales last month are not always the same as in the future, so external data is needed in predicting future sales.

Therefore, in this thesis a prediction will be made using 3 models, namely the GRU, SVR and SARIMAX models with the help of external data in the form of CPI data and inflation data. In addition, this thesis also conducted a correlation test to determine whether the sales data to be predicted has significance/relationship with external data so that it helps in predicting sales data.

*The results obtained from this study are that *pot* data is more suitable for using univariate data with the GRU model, with RMSE Train 3.22, RMSE Test 2.93. For *hanger* and *sealware* data, the best model for prediction is the SARIMAX model with univariate data type (RMSE 30.43) and multivariate data type (RMSE 8.07).*

Keywords: *Sales Forecasting, GRU, SARIMAX, SVR, CPI, Inflation, sales prediction*

1. PENDAHULUAN

Salah satu hal yang penting untuk dimiliki sebuah perusahaan yang bergerak di bidang retail maupun distributor adalah sebuah sistem *sales forecasting*. Sebuah *sales forecast* dinyatakan penting karena sangat berdampak bagi perusahaan, dimana *forecast* yang buruk dapat berdampak pada *inventory* perusahaan sehingga dapat menyebabkan penyimpanan barang terlalu banyak maupun terlalu sedikit sehingga menyebabkan perusahaan rugi [8]. Selain itu, *sales forecast* yang buruk juga dapat membuat beberapa resources terbuang karena kecenderungan customer yang tidak ingin barang terlalu lama disimpan [2]. Permasalahan penyimpanan barang yang tidak tepat, serta permintaan customer mengarah pada produk terlalu lama disimpan telah beberapa kali dialami oleh perusahaan X, dimana perusahaan X merupakan sebuah perusahaan yang bergerak di bidang distributor plastik. Masalah yang dialami perusahaan juga disebabkan adanya perubahan *trend* yang terjadi secara tidak terduga sehingga diperlukan sebuah analisa yang tidak hanya mempertimbangkan data penjualan sebelumnya, tetapi perlu dilakukan pengujian data – data external yang dapat berpengaruh pada penjualan perusahaan, seperti data inflasi dan data IHK. Pada penelitian ini, *sales forecasting* akan dilakukan dengan menggunakan data perusahaan X untuk melakukan prediksi jumlah penjualan untuk produk dalam jangka waktu 6 bulan ke depan, dan prediksi dilakukan supaya perusahaan dapat menyesuaikan pembelian produk dengan melihat hasil *forecast* ke depannya sehingga perusahaan bisa menyimpan stock secukupnya, dapat melihat laju penjualan suatu produk untuk beberapa waktu ke depan.

Agar perusahaan dapat mencapai hal tersebut, perlu dilakukan sebuah analisa yang juga melibatkan data – data external yaitu data IHK dan Inflasi. Pada pemilihan data external inflasi didasarkan

pada penelitian yang melakukan analisa pengaruh inflasi terhadap pertumbuhan industri pengolahan di Indonesia [10]. Penelitian ini menjelaskan kenaikan inflasi memiliki pengaruh yang buruk terhadap pertumbuhan industri pengolahan di Indonesia karena kenaikan inflasi akan mengakibatkan kenaikan harga barang yang menyebabkan kemampuan beli masyarakat menurun, sehingga secara tidak langsung berdampak pada penurunan pada penjualan di sektor industri pengolahan [10]. Dari beberapa pengujian yang dilakukan, h1 diterima sehingga bisa disimpulkan inflasi berpengaruh pada pertumbuhan industri pengolahan [10]. Lalu, data IHK dipilih berdasarkan penelitian mengenai analisa pengaruh indeks harga konsumen terhadap tingkat konsumsi rumah tangga di Provinsi Banten. Penelitian ini menjelaskan indeks harga konsumen merupakan sebuah indeks untuk mengukur perubahan harga barang dan jasa yang dikonsumsi oleh rumah tangga atau masyarakat dalam kurun waktu tertentu, dan salah satu penentu tingkat konsumsi barang rumah tangga adalah harga barang. Hasil yang diperoleh oleh penelitian ini adalah hipotesa 1 diterima, dimana IHK berpengaruh signifikan pada tingkat konsumsi rumah tangga. Kedua data external nantinya akan dimasukkan pada analisa saat melakukan prediksi dengan model dan akan diuji apakah data external memiliki pengaruh pada hasil forecast yang dihasilkan / tidak.

Oleh karena itu, pada penelitian ini akan dilakukan prediksi pada data penjualan dengan menggunakan 3 model, yaitu model SVR, SARIMAX dan GRU. Prediksi yang dilakukan juga akan menggunakan bantuan data external, yaitu data Inflasi dan IHK. Penambahan data external dilakukan dengan harapan dapat menghasilkan akurasi yang lebih baik daripada hanya menggunakan data penjualan saja. Selain itu, diharapkan hasil penelitian ini dapat membantu perusahaan untuk mengetahui data mana yang memiliki keterkaitan dengan data Inflasi dan IHK.

2. LANDASAN TEORI

2.1 Tinjauan Pustaka

Metode yang digunakan dalam penelitian ini terbagi menjadi 3, yaitu SARIMAX, GRU dan SVR. Berikut di bawah ini akan dijelaskan sekilas mengenai metode yang digunakan.

2.1.1 SVR

SVM merupakan metode yang dikembangkan oleh Vapnik dan pertama kali SVM dikembangkan untuk masalah classification dan regression. Lalu, Drucker et al. mengembangkan model yang bernama SVR berdasarkan konsep dari model SVM dan SVR ditujukan untuk mengurangi error rate dengan meminimalkan jarak antara hasil yang diprediksi dan data training [10]. SVR yang merupakan pengembangan dari SVM saat ini digunakan pada permasalahan pattern recognition dan data analysis menggunakan regression [4]. Sebelum membahas lebih lanjut mengenai rumus dasar SVR, perlu diketahui equation dasar berikut ini, yaitu $S = [(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_N, y_N)]$, dimana x_i merupakan input data n dimension, y_i merupakan output variable, dan N merupakan total data [11]. Dengan menggunakan SVR, untuk melakukan estimasi output y_i dengan independent variable yaitu x_i , maka disusunlah sebuah rumus dasar seperti pada Rumus nomor 1.

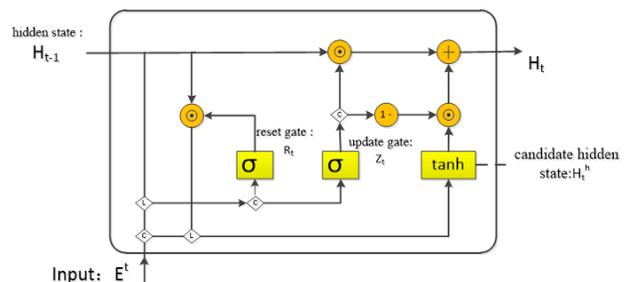
$$f_{(x)} = \omega^T x + b \quad (1)$$

2.1.2 SARIMAX

SARIMAX merupakan sebuah model yang diambil dari konsep pendahulunya, yaitu dari model SARIMA dan ARIMA. Pada dasarnya, model ARIMA pada time series memiliki 3 variabel utama, yaitu p sebagai autoregressive, q sebagai moving average dan d sebagai non-seasonal difference [1]. Model SARIMA sendiri memiliki konsep yang serupa dengan model ARIMA, dimana perbedaannya hanya terletak pada tambahan elemen *seasonality*.

2.1.3 GRU

GRU merupakan sebuah model yang hampir serupa dengan LSTM dengan struktur yang lebih sederhana bila dibandingkan dengan LSTM [3], dimana model GRU memiliki reset gate dan update gate. Berikut pada Gambar 1 dapat dilihat secara sekilas struktur dari model GRU.



Gambar 1. Struktur Model GRU berupa update dan reset gate

Reset gate memiliki tugas untuk memilah seberapa banyak data masa lampau yang digunakan dan menghapus data yang tidak diperlukan, sedangkan update gate menentukan update rate dalam mengolah data lama dan baru [5]. Dengan begitu dapat disimpulkan secara tidak langsung update gate dan reset gate dapat menyelesaikan permasalahan yang seringkali dialami oleh model RNN, yaitu vanishing gradient problem, dimana gate dapat di train untuk mempertahankan data yang lama dan menggabungkannya dengan yang baru [7].

2.2 Tinjauan Studi

Pada penelitian ini akan menggunakan beberapa studi yang telah dikembangkan oleh beberapa peneliti berkaitan dengan metode beserta dengan pemilihan data external.

Salah satu *studi* dikembangkan untuk menjawab sebuah masalah, yaitu sebuah perusahaan Y yang ingin mengurangi biaya produksi dengan cara membeli secara langsung dari produsen sehingga dapat meningkatkan demand customer [6]. Dari segi hasil yang dihasilkan pada *studi* ini menunjukkan bahwa SARIMA model dapat melakukan prediksi pada data yang mengandung *seasonality* dengan menghasilkan MAD value berkisar 5.592 yang merupakan score terbaik [6].

Lalu, pemilihan data IHK sebagai data external didasarkan pada sebuah *studi* yang digagas oleh peneliti bernama Sari, dimana peneliti melakukan pengujian hipotesis yang digagas mengenai pengaruh IHK dan Pendapatan per kapita terhadap tingkat konsumsi masyarakat di kota Banjarmasin [9]. Dari segi hasil studi, peneliti menyimpulkan bahwa IHK dan Pendapatan per kapita sangat signifikan mempengaruhi tingkat konsumsi di Banjarmasin secara bersamaan / simultan, dimana peneliti menjelaskan bahwa

kenaikan harga yang salah satunya karena IHK naik membuat konsumen menyimpan uang untuk membeli sesuai yang mereka miliki, sehingga berpengaruh pada tingkat konsumsi yang dilakukan [9].

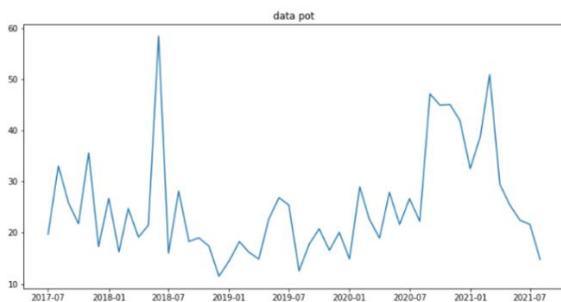
Setelah itu, pemilihan data Inflasi sebagai data external didasarkan pada *studi* yang dilakukan oleh Rulinawati tahun 2017, dimana masalah yang diangkat yaitu pengujian pada hipotesis digagas mengenai pengaruh inflasi terhadap pertumbuhan industri pengolahan di Indonesia. Hal ini bertujuan untuk mengetahui sejauh mana pengaruh yang dihasilkan inflasi terhadap pertumbuhan industri pengolahan. Dari hasil pengujian, diduga H_1 yang diusulkan oleh peneliti diterima sehingga bisa disimpulkan bahwa inflasi berpengaruh terhadap pertumbuhan industri pengolahan di Indonesia. Lalu, perbedaan penelitian yang dilakukan dengan skripsi ini adalah penelitian ini melakukan pengujian hipotesis yang mengarah pada pengaruh inflasi terhadap pertumbuhan industri pengolahan secara general yang didalamnya terdapat berbagai macam produk seperti makanan, minuman, mesin, dan lain – lain.

3. DESAIN SISTEM

Dataset yang digunakan dalam penelitian ini adalah data penjualan perusahaan yang dipilih berdasarkan keterkaitan dengan data Inflasi dan data IHK, yaitu data pot, data hanger dan data sealware.

3.1 Data Pot

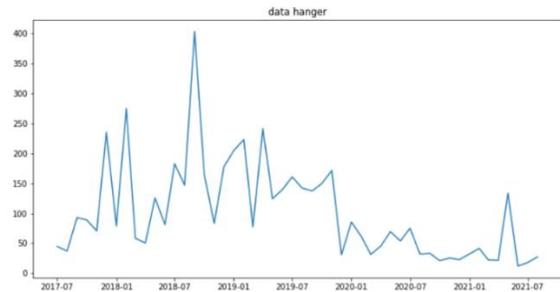
Pada Gambar 2 yang menunjukkan data pot selama 4 tahun, ada beberapa poin penting yang bisa diambil setelah melihat data di atas, salah satunya adalah kenaikan penjualan pot pada bulan 2018, dimana tahun 2018 merupakan tahun kabisat dan kenaikan penjualan pot memuncak dengan rata – rata penjualan mencapai 58.4 pada bulan 6 tahun 2018 bersamaan dengan hari raya lebaran dan idul fitri. Lalu, terjadi kenaikan dan penurunan penjualan berturut – turut sebanyak 3 kali pada penjualan pot yang dimulai saat bulan oktober 2017 hingga bulan april 2018 sebelum akhirnya penjualan pot memuncak, dan hal ini tampak terulang kembali pada januari tahun 2020 hingga bulan agustus 2020 terjadi kenaikan dan penurunan penjualan sebanyak 3 kali sebelum akhirnya penjualan pot naik drastis. Selain itu, kenaikan puncak penjualan yang sebelumnya terjadi pada tahun 2018 bulan 6 dengan rata – rata penjualan 58.4 juga terulang kembali pada 2021 bulan 3 dengan rata – rata penjualan 50.94 yang berarti kemungkinan penjualan memuncak sekitar 2 setengah tahun sampai 3 tahun, dan setelah memuncak akan membutuhkan 5 – 6 bulan untuk penjualan turun dengan rata – rata penjualan 10 – 20.



Gambar 2. Gambar Plot Data Pot

3.2 Data Hanger

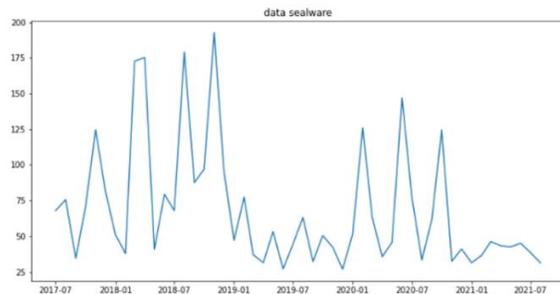
Pada Gambar 3 yang menunjukkan data hanger selama 4 tahun, ada beberapa poin penting yang bisa diambil setelah melihat data di atas, salah satunya yaitu pola laju penjualan yang terhenti pada saat memasuki tahun 2020 bulan februari, dimana saat itu terjadi pandemic covid 19 yang membuat penjualan hanger menurun drastis dari tahun – tahun sebelumnya. Hal ini kemungkinan disebabkan karena bulan 5 tahun 2021 yang bertepatan dengan hari raya lebaran dan idul fitri. Lalu, terdapat *cycle* penjualan pada akhir tahun 2017 – awal tahun 2018 dengan akhir tahun 2018 – awal tahun 2019, tetapi *cycle* hanya berhenti sampai di waktu tersebut.



Gambar 3. Hasil Plot Data Hanger

3.3 Data Sealware

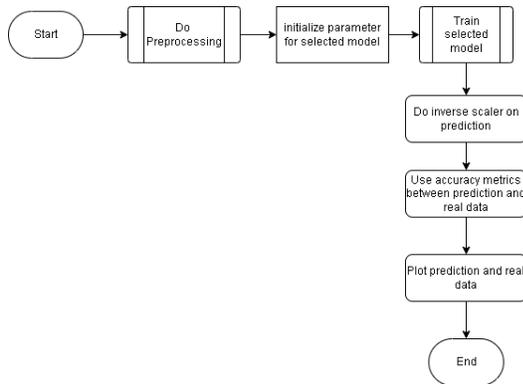
Pada Gambar 4 yang menunjukkan data sealware selama 4 tahun, ada beberapa poin penting yang bisa diambil setelah melihat data di atas, salah satunya yaitu penjualan yang cenderung naik pada tahun kabisat, yaitu tahun 2018 dan 2020, khususnya pada saat tahun 2018 bulan 11 serta tahun 2020 bulan 10 mengalami kenaikan penjualan dengan rata – rata di atas 100 (tahun 2018 bulan 11 dan tahun 2020 bulan 10 bertepatan dengan Maulid Nabi Muhammad SW).



Gambar 4. Hasil Data Plot Sealware

3.4 Analisa Sistem

Pada subbab ini akan dibahas proses *flowchart* train test pada sistem yang digunakan pada penelitian ini, beserta dengan proses *flowchart* untuk data *preprocessing*.



Gambar 5. Flowchart proses train test model

Berdasarkan Gambar 5, berikut merupakan proses *flowchart* train test model pada penelitian ini, dimana proses train test model dimulai dengan adanya proses *preprocessing data* dan dilanjutkan dengan inialisasi parameter model yang dipilih. Proses *preprocessing data* terdiri atas pembacaan data penjualan bentuk file .csv, melakukan *summary*, melakukan *join* dataset menjadi satu table, mencari korelasi antara data penjualan dengan data external dan diakhiri dengan melakukan normalisasi pada data. Lalu, model yang sudah dilakukan inialisasi akan dilakukan train menggunakan data penjualan perusahaan beserta dengan data external, dan setelah itu proses dilanjutkan dengan *inverse scaler* untuk membuang nilai normalisasi. Langkah selanjutnya adalah mencari akurasi pada metode akurasi pada hasil prediksi dengan data asli, dan diakhiri dengan melakukan plot hasil prediksi.

4. PENGUJIAN

Pengujian akan dilakukan dalam beberapa tahap. Pertama akan dilakukan uji korelasi antara data external dengan data penjualan perusahaan, dan yang kedua adalah pengujian hasil prediksi model pada tiap – tiap data penjualan yang digunakan.

4.1 Pengujian Uji Korelasi

Pengujian hasil uji korelasi antara data penjualan perusahaan X dengan data external dilakukan sebelum melakukan training model, dimana hal ini dilakukan untuk mencari tahu apakah data penjualan barang plastik memiliki keterkaitan / signifikansi dengan data external yaitu data Inflasi dan IHK. Pengujian korelasi dilakukan menggunakan library Pandas dengan menggunakan function `series.corr`, dan data penjualan yang dipakai saat melakukan pengujian dibagi menjadi beberapa kategori barang, di antaranya yaitu general, sealware/toples, mkn, dpr, pot, hngr, tpm, btlair, bak/ember, dan disp. Tujuan penggunaan shifted pada data external adalah supaya menghasilkan analisis yang memperhitungkan relevansi data bulan sebelumnya sehingga bisa diketahui korelasi antara data external dengan data penjualan pada bulan yang sama maupun terdapat selisih bulan yang berbeda.

Dari hasil uji korelasi pada Tabel 1, sebagian besar data penjualan perusahaan tidak memiliki signifikansi / keterkaitan dengan data Inflasi dan IHK, dan hal ini bisa dilihat dari nilai antara kategori barang dengan kolom `ihk_0` sampai `inflasi_2` yang berkisar antara 0.006 sampai dengan 0.38. Data hanger memiliki korelasi yang kuat dengan data IHK karena nilai korelasi melebihi 0.5 untuk data IHK, data IHK shift 1 dan data IHK shift 2, sehingga data hngr memiliki indikasi berkorelasi positif dengan data IHK. Data lainnya yang terlihat memiliki indikasi korelasi kuat adalah data general yang berkorelasi negatif dengan data Inflasi yang bernilai -0.42.

4.2 Pengujian Hasil Uji Korelasi dan Prediksi Model

Berdasarkan Tabel 2, terdapat indikasi bahwa model SVR pada data pot memiliki akurasi RMSE dan MAE yang paling buruk apabila dibandingkan dengan model yang lain (tipe data multivariate dan multivariate shift 1 menghasilkan akurasi RMSE test 10.44 dan 13.13, beserta dengan MAE test yaitu 7.59 dan 8.46). Model SVR pada data pot cenderung kurang stabil dikarenakan prediksi model SVR cenderung memiliki error rate yang tinggi saat melakukan prediksi pada penjualan dengan kenaikan yang besar, sehingga secara tidak langsung berpengaruh dengan tingginya nilai RMSE dan MAE pada model SVR.

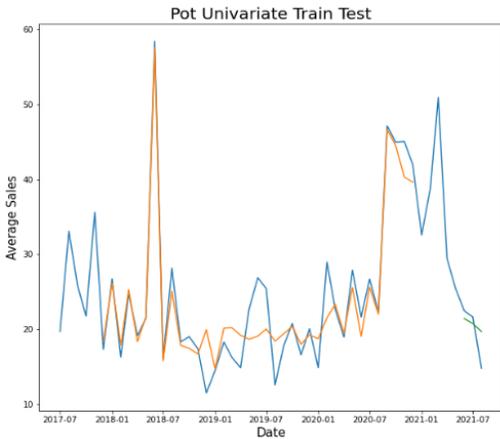
Tabel 1. Tabel Hasil Uji Korelasi

	Ihk_0	Ihk_1	Ihk_2	Inflasi_0	Inflasi_1	Inflasi_2
General	-0.179	-0.214	-0.261	-0.179	-0.420	-0.191
Sealware/toples	0.145	0.129	0.197	-0.040	-0.032	0.033
Peralatan makan	0.058	0.114	0.188	0.243	0.151	-0.025
Peralatan dapur	-0.132	-0.153	-0.098	-0.138	-0.247	-0.307
Pot	-0.403	-0.453	-0.448	0.021	-0.107	-0.113
Hanger	0.604	0.609	0.598	0.114	0.178	0.229
tpm	-0.255	-0.275	-0.286	-0.075	-0.053	-0.014
Btlair	-0.136	-0.138	-0.141	-0.051	-0.006	0.076
Bak/ember	0.257	0.230	0.215	0.056	-0.031	-0.120
Dispenser	-0.330	-0.381	-0.254	-0.062	-0.026	0.032

Tabel 2. Tabel Pengujian Data Pot

Model	RMSE train	MAE train	RMSE Test	MAE test
SVR Multivariate	8.78	5.64	10.44	7.59
SVR Multivariate shift 1	8.98	5.41	13.13	8.46
SARIMAX Univariate	-	-	7.35	4.56
SARIMAX Multivariate	-	-	9.1	7.31
SARIMAX Multivariate shift 1	-	-	8.30	7.31
GRU Univariate	3.22	2.28	2.93	2.26
GRU Multivariate	2.23	1.6	5.98	3.72
GRU Multivariate shift 1	2.96	2.26	4.4	3.55

Selain itu, pada Tabel 2 untuk data pot pada model SARIMAX dan GRU, terdapat indikasi bahwa model GRU memiliki akurasi yang lebih baik daripada model SARIMAX apabila dibandingkan menggunakan akurasi RMSE maupun MAE. Hal ini dapat dilihat pada hasil akurasi RMSE dan MAE test yang dihasilkan GRU tipe data univariate dengan nilai 2.93 dan 2.26 yang lebih kecil daripada akurasi RMSE dan MAE test yang dihasilkan oleh model SARIMAX tipe data *univariate* dengan nilai 7.35 dan 4.56. Selain akurasi, terdapat indikasi lain yang nampak dari Gambar 6 yaitu tipe data yang cocok untuk data pot adalah tipe data univariate, karena pada tipe data *univariate* model GRU menghasilkan prediksi yang cenderung sama dengan data asli.



Gambar 6. Hasil prediksi data pot univariate model GRU

Tabel 3. Tabel Pengujian Data Hanger

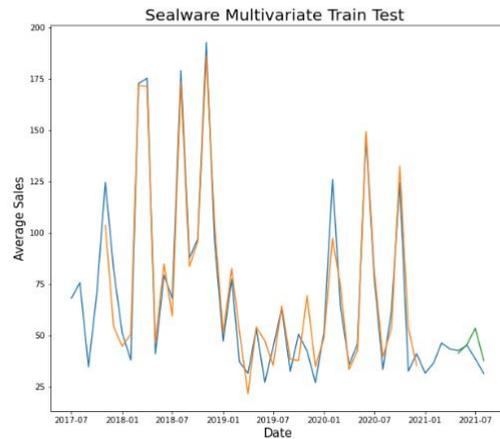
Model	RMSE train	MAE train	RMSE Test	MAE test
SVR <i>Multivariate</i>	68.15	43.67	69.47	68.35
SVR <i>Multivariate shift 1</i>	71.03	45.07	67.65	67.3
SARIMAX <i>Univariate</i>	-	-	30.43	23.0
SARIMAX <i>Multivariate</i>	-	-	30.88	20.31
SARIMAX <i>Multivariate shift 1</i>	-	-	31.6	24.72
GRU <i>Univariate</i>	46.09	35.62	56.52	39.87
GRU <i>Multivariate</i>	46.6	33.39	45.5	30.32
GRU <i>Multivariate shift 1</i>	56.97	39.72	49.56	33.9

Berdasarkan hasil akurasi pada Tabel 3 dan Tabel 4 untuk data hanger serta data sealware, terdapat indikasi bahwa hasil prediksi yang dihasilkan oleh model SVR memiliki hasil akurasi yang paling kurang apabila dibandingkan dengan model lainnya sama seperti dengan hasil prediksi pada data pot. Hasil akurasi untuk tipe data multivariate dan multivariate shift 1 menghasilkan akurasi RMSE test 69.47 dan 68.35, beserta dengan MAE test yaitu 67.65 dan 67.3 pada data hanger, sedangkan untuk data sealware

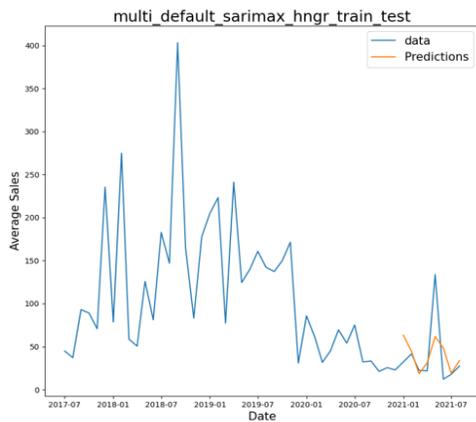
menghasilkan nilai RMSE test 9.32 dan 8.28, beserta dengan MAE test yaitu 12.89 dan 12.03. Prediksi model SVR pada data hanger cenderung memiliki indikasi yang sama dengan prediksi pada data pot, dimana model memiliki error rate yang tinggi saat melakukan prediksi pada penjualan dengan kenaikan yang besar.

Tabel 4. Tabel Pengujian Data Sealware

Model	RMSE train	MAE train	RMSE Test	MAE test
SVR <i>Multivariate</i>	33.59	20.44	9.32	8.28
SVR <i>Multivariate shift 1</i>	39.84	23.29	12.89	12.03
SARIMAX <i>Univariate</i>	-	-	8.26	7.09
SARIMAX <i>Multivariate</i>	-	-	8.07	6.55
SARIMAX <i>Multivariate shift 1</i>	-	-	6.53	4.95
GRU <i>Univariate</i>	17.97	14.38	13.82	12.72
GRU <i>Multivariate</i>	11.69	8.93	8.07	5.69
GRU <i>Multivariate shift 1</i>	38.7	28.9	9.88	7.56



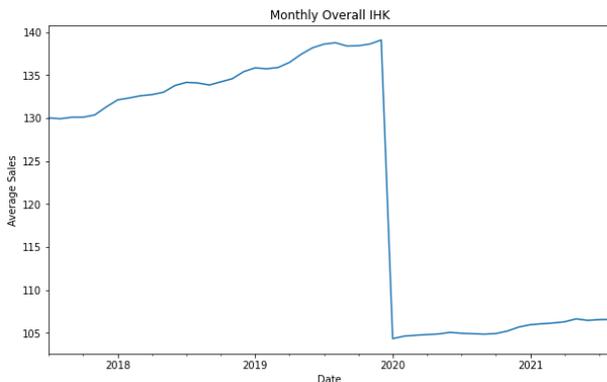
Gambar 7. Hasil prediksi data sealware multivariate model GRU



Gambar 8. Hasil prediksi data hanger multivariate model SARIMAX

Berdasarkan hasil akurasi pada Tabel 3 dan Gambar 8 untuk plot prediksi untuk data hanger pada model SARIMAX dan GRU, terdapat indikasi bahwa model SARIMAX memiliki nilai akurasi RMSE maupun MSE lebih baik daripada model GRU. Dari segi nilai akurasi yang dihasilkan model SARIMAX pada tipe data univariate, terdapat nilai RMSE test 30.43, MAE test 23 maupun multivariate dengan nilai RMSE test 30.88, MAE test 20.31 yang lebih kecil daripada akurasi RMSE dan MAE test yang dihasilkan oleh model GRU tipe data univariate dengan nilai RMSE test 56.52, MAE test 39.87 dan tipe data multivariate dengan nilai RMSE test 45.5, MAE test 30.32. Selain itu, terdapat indikasi bahwa tipe data univariate pada data hanger cenderung mengalami *overfitting* dimana model GRU memiliki prediksi yang hampir akurat pada data train, tetapi tidak terlihat akurat saat prediksi pada data test.

Berdasarkan hasil akurasi pada Tabel 4 dan Gambar 7 untuk plot prediksi untuk data sealware pada model SARIMAX dan GRU, terdapat indikasi bahwa model GRU dan SARIMAX memiliki akurasi yang hampir serupa, khususnya saat menggunakan akurasi RMSE pada pengujian akurasi. Hal ini dapat dilihat pada hasil akurasi RMSE dan MAE test yang dihasilkan model GRU tipe data multivariate dengan nilai 8.07 serta 5.69, dimana nilai RMSE test ini memiliki kesamaan dengan model SARIMAX dan nilai MAE test model SARIMAX yaitu 6.55.



Gambar 9. Gambar plot data IHK

Bila dilihat dari hasil prediksi data sealware dan hasil dari uji korelasi pada Tabel 1, hasil korelasi mengindikasikan bahwa data sealware tidak memiliki korelasi yang kuat dengan kedua data

external, tetapi pada kenyataannya prediksi data sealware dengan tipe data multivariate memiliki hasil yang lebih bagus bila dibandingkan dengan tipe data univariate. Hal ini mengindikasikan bahwa korelasi yang lemah tidak selalu menyebabkan akurasi pada jenis data univariate lebih tinggi, karena korelasi hanya membantu untuk menentukan data mana yang memiliki keterkaitan tetapi tidak menentukan hasil akhir. Selain itu, pada data hanger hasil prediksi model SARIMAX memiliki akurasi yang lebih baik secara keseluruhan karena diduga model SARIMAX memiliki performa yang lebih baik saat digunakan pada data yang tidak terlalu banyak, dimana pada penelitian ini data yang digunakan hanya berjumlah 50.

Pengujian uji korelasi yang dilakukan pada poin 4.1 tidak semata dilakukan untuk mencari keterkaitan antara data penjualan perusahaan dengan data external, tetapi terdapat dugaan yaitu kenaikan / penurunan harga yang diduga dapat mempengaruhi jumlah penjualan perusahaan, dimana ketika harga barang naik maka penjualan turun dan sebaliknya. Berdasarkan Tabel 1, terdapat indikasi bahwa data hanger memiliki korelasi positif yang kuat dengan data IHK dengan nilai 0.6, dan hal ini terlihat pada Gambar 9 data IHK tahun 2020 mengalami penurunan yang signifikan beserta dengan penjualan data hanger pada Gambar 3 sehingga arah kedua data di atas sama. Hasil korelasi data hanger dan data IHK mengindikasikan bahwa laju penjualan perusahaan tidak selalu dipengaruhi oleh naik / turunnya harga, tetapi lebih dipengaruhi oleh permintaan customer pada suatu kondisi tertentu dan juga pada musim tertentu (tahun 2020 awal pandemi covid 19 terdapat penurunan yang signifikan pada data IHK bersamaan dengan turunnya permintaan customer pada jenis barang hanger).

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian pada sistem, dapat disimpulkan sebagai berikut:

- Pada beberapa data penjualan seperti data sealware dan hanger, metode SARIMAX memiliki performa yang lebih baik daripada model GRU karena diduga jumlah data saat training mempengaruhi kinerja model yang menyebabkan model SARIMAX menjadi lebih bagus daripada model GRU
- Berdasarkan hasil RMSE dan analisa hasil prediksi, jenis data multivariate dan univariate sangat cocok digunakan pada model GRU dan SARIMAX pada data penjualan sealware dan hanger, sedangkan jenis data univariate model GRU lebih baik digunakan pada penjualan pot
- Hasil uji korelasi mengindikasikan bahwa tingkat penjualan perusahaan tidak selalu dipengaruhi oleh naik / turunnya harga jual barang, tetapi lebih dipengaruhi oleh permintaan / demand customer pada kurun waktu tertentu / sesuai dengan musim – musim tertentu
- Korelasi yang lemah tidak selalu menyebabkan akurasi pada jenis data univariate lebih tinggi, karena korelasi hanya menentukan data mana yang memiliki keterkaitan (tidak menentukan hasil akhir)
- Hasil uji korelasi yang dilakukan mengindikasikan bahwa data hanger dan data pot memiliki keterkaitan yang kuat dengan data external IHK (nilai korelasi 0.6 dan -0.4)

Saran

- Kondisi data harus stabil ketika melakukan training data untuk menghindari bias (*overfit* pada data) yang disebabkan oleh perubahan pola data secara tiba-tiba

6. DAFTAR PUSTAKA

- [1] Arunraj, N. S., Ahrens, D., & Fernandes, M. 2016. Application of SARIMAX model to forecast daily sales in food retail industry. *International Journal of Operations Research and Information Systems (IJORIS)*, 7(2), 1-21. DOI = 10.4018/IJORIS.2016040101
- [2] Bohanec, M., Robnik-Šikonja, M., & Borštnar, M. K. 2017. Organizational learning supported by machine learning models coupled with general explanation methods: A Case of B2B sales forecasting. *Organizacija*, 50(3), 217-233. DOI = 10.1515/orga-2017-0020
- [3] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv = 1412.3555*.
- [4] Meighani, H. M., Ghotbi, C., Behbahani, T. J., & Sharifi, K. 2018. Evaluation of PC-SAFT model and Support Vector Regression (SVR) approach in prediction of asphaltene precipitation using the titration data. *Fluid Phase Equilibria*, 456, 171-183. DOI = 10.1016/j.fluid.2017.10.022
- [5] Park, H., & Hwang, S. 2021. Demand forecasting of micro mobility using a gated recurrent unit. DOI = 10.22712/susb.20210014
- [6] Pongdatu, G. A. N., & Putra, Y. H. 2018, August. Seasonal time series forecasting using SARIMA and Holt Winter's exponential smoothing. In *IOP Conference Series: Materials Science and Engineering* (Vol. 407, No. 1, p. 012153). IOP Publishing. DOI = 10.1088/1757-899X/407/1/012153
- [7] Rahman, M. O., Hossain, M. S., Junaid, T. S., Forhad, M. S. A., & Hossen, M. K. 2019. Predicting prices of stock market using gated recurrent units (GRUs) neural networks. *Int. J. Comput. Sci. Netw. Secur*, 19(1), 213-222. DOI = 10.17577/IJERTV9IS050550
- [8] Ramos, P., Santos, N., & Rebelo, R. 2015. Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing*, 34, 151-163. DOI = 10.1016/j.rcim.2014.12.015
- [9] Sari, N. Z. A. 2019. Pengaruh Pendapatan Perkapita dan IHK Terhadap Konsumsi Masyarakat di Kota Banjarmasin. *JIEP: Jurnal Ilmu Ekonomi dan Pembangunan*, 1(1). DOI = 10.20527/jiep.v1i1.1129
- [10] Singh, A., Kotiyal, V., Sharma, S., Nagar, J., & Lee, C. C. 2020. A machine learning approach to predict the average localization error with applications to wireless sensor networks. *IEEE Access*, 8, 208253-208263.
- [11] Singh, A., Gaurav, K., Rai, A. K., & Beg, Z. 2021. Machine learning to estimate surface roughness from satellite images. *Remote Sensing*, 13(19), 3794. DOI = 10.3390/rs13193794