

Perbandingan Metode *Machine Learning* Dalam Peramalan Kebutuhan Semen

Andi Muhammad Fiqri Achmad¹, Iksan Adiasa²
^{1,2} Politeknik ATI Makassar
am_fiqri@atim.ac.id¹, iksan.adiasa@atim.ac.id²

ABSTRAK

PT. X adalah salah satu perusahaan penyedia beton ready mix untuk prasarana jalan beton. Pesanan beton yang masuk ke perusahaan hampir setiap hari dengan jumlah yang tidak tentu. Hal itu membuat perusahaan membutuhkan banyak semen sebagai bahan baku beton ready mix. Apabila semen yang disimpan di silo tidak cukup, maka perusahaan harus memesan dahulu semen curah ke perusahaan mitra yang tentunya membutuhkan waktu beberapa jam. Akibatnya pesanan pengguna dapat tertunda juga. Penelitian ini bertujuan untuk membandingkan hasil peramalan kebutuhan semen dari beberapa algoritma machine learning. Penelitian dimulai dari mengidentifikasi gejala masalah empiris dan studi literatur terkait penelitian – penelitian terkait metode peramalan dan peramalan kebutuhan semen. Selanjutnya dilakukan input data dan rekayasa data dimana variabel waktu dibagi menjadi beberapa feature, yaitu dayofweek, quarter, month, dayofyear, dayofmonth, dan weekofyear, dilakukan data scaling dan uji cross-validation. Beberapa algoritma machine learning digunakan untuk peramalan, antara lain Support Vector Regression, Linear Regression, Random Forest Regressor, Decision Tree Regressor, Ridge, Lasso, Elastic Net, Extreme Gradient Boosting (XGB), dan XGBRF. Diperoleh bahwa algoritma Ridge memiliki nilai kesalahan yang paling kecil diantara algoritma lainnya setelah dilakukan uji cross-validation dengan nilai RMSE masing – masing sebesar 0,153861, 0,171343, dan nilai MAE sebesar 0,114138.

Kata kunci: peramalan, semen, *machine learning*

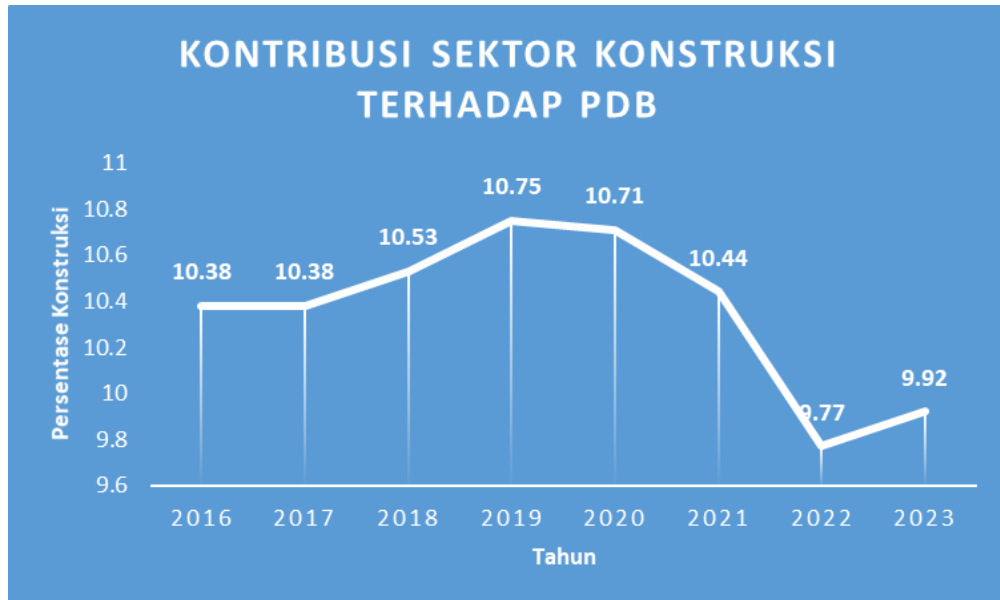
ABSTRACT

PT. X is one of the companies providing ready-mix concrete for concrete road infrastructure. Concrete orders that come into the company almost every day with uncertain amounts. This makes the company need a lot of cement as a raw material for ready-mix concrete. If the cement stored in the silo is not enough, the company must first order bulk cement from a partner company which of course takes several hours. As a result, user orders can also be delayed. This study aims to compare the results of cement demand forecasting from several machine learning algorithms. The study began by identifying empirical problem symptoms and literature studies related to research - research related to forecasting methods and forecasting cement needs. Furthermore, data input and data engineering were carried out where the time variable was divided into several features, namely dayofweek, quarter, month, dayofyear, dayofmonth, and weekofyear, data scaling and cross-validation tests were carried out. Several machine learning algorithms are used for forecasting, including Support Vector Regression, Linear Regression, Random Forest Regressor, Decision Tree Regressor, Ridge, Lasso, Elastic Net, Extreme Gradient Boosting (XGB), and XGBRF. It was found that the Ridge algorithm had the smallest error value among the other algorithms after cross-validation testing was carried out with RMSE values of 0.153861, 0.171343, and MAE values of 0.114138.

Keywords: forecasting, cement, machine learning

PENDAHULUAN

Konstruksi merupakan salah satu sektor yang berperan penting terhadap pembangunan nasional. Berdasarkan Informasi Statistik Infrastruktur PUPR 2022 dan 2023, rata – rata kontribusi sektor konstruksi tahun 2016 - 2023 yaitu sebesar 10,36%, dengan nilai tertinggi pada tahun 2019 dan terendah pada tahun 2023 [1]. Tingginya kontribusi ini juga turut serta mendorong peningkatan kebutuhan pembangunan, baik dari sisi material, teknologi, dan sumber daya manusia. Hal ini menyebabkan perencanaan kebutuhan tersebut harus dilakukan dengan tepat.



Gambar 1. Kontribusi sektor konstruksi terhadap pendapatan domestic bruto

Jalan adalah salah satu objek konstruksi yang paling banyak dibangun, baik jalan nasional, jalan daerah, maupun jalan tol. Jalan yang dibangun terdiri dari dua jenis, yaitu jalan aspal dan jalan beton [2]. Dalam membangun jalan beton, salah satu material utamanya adalah semen. Semen digunakan sebagai bahan perekat dan bahan campuran beton [3]. Selanjutnya, semen akan dicampurkan dengan bahan–bahan kimia lainnya untuk memperkuat beton yang akan dicetak. Kualitas semen yang digunakan akan sangat berpengaruh terhadap daya tahan dan kekuatan beton yang dihasilkan, sehingga pemilihan semen yang sesuai dengan standar teknis menjadi hal yang sangat penting. Selain itu, proses pencampuran beton juga harus memperhatikan proporsi bahan yang tepat agar hasil akhirnya memiliki ketahanan terhadap beban lalu lintas dan perubahan cuaca.

PT. X adalah salah satu perusahaan penyedia beton ready mix, khususnya untuk prasarana jalan beton. Pesanan beton yang masuk ke perusahaan hampir setiap hari dengan jumlah yang tidak tentu. Kadang dalam satu hari terdapat beberapa pesanan yang masuk sekaligus. Hal itu membuat perusahaan membutuhkan banyak semen sebagai bahan baku beton ready mix. Apabila semen yang disimpan di silo tidak cukup, maka perusahaan harus memesan dahulu semen curah ke perusahaan mitra yang tentunya membutuhkan waktu beberapa jam. Akibatnya, pesanan pengguna dapat tertunda juga, dan dapat mengakibatkan penurunan kepuasan konsumen dari segi waktu. Selain itu, keterlambatan pasokan semen juga berpotensi mengganggu efisiensi operasional perusahaan, karena dapat menyebabkan mesin–mesin produksi tidak berjalan optimal dan tenaga kerja menjadi tidak produktif selama menunggu bahan baku tersedia. Oleh karena itu, dibutuhkan perencanaan yang tepat dengan melakukan peramalan jumlah semen yang dibutuhkan tiap hari. Dengan peramalan yang akurat, perusahaan dapat mengelola persediaan semen dengan lebih optimal, mengurangi risiko keterlambatan produksi, serta meningkatkan kepuasan pelanggan melalui pemenuhan pesanan yang lebih cepat dan efisien.

Metode peramalan *time series* telah berkembang dengan pesat, dari metode statistika hingga penggunaan algoritma *machine learning* [4]. Penelitian menggunakan metode statistika, seperti *moving average*, *exponential smoothing*, *autoregressive integrated moving average* (ARIMA) biasanya digunakan sebagai pilihan awal dalam melakukan peramalan. Selanjutnya terdapat juga metode Prophet yang dikembangkan oleh Facebook atau Meta, dan sudah cukup banyak digunakan [5]. Sedangkan algoritma *machine learning* sendiri mulai cukup sering digunakan, baik algoritma *Support Vector Regression* [6], *Decision Tree* [7], *Naïve Bayes* [8], dan sebagainya. Karena banyaknya metode yang dapat digunakan untuk melakukan peramalan, sehingga setiap metode tersebut sebaiknya dibandingkan terlebih dahulu sebelum diterapkan. Oleh

karena itu, penelitian ini bertujuan untuk membandingkan hasil peramalan kebutuhan semen dari beberapa algoritma *machine learning*. Batasan masalah dalam penelitian ini yaitu hanya menggunakan satu objek yaitu semen.

METODE PENELITIAN

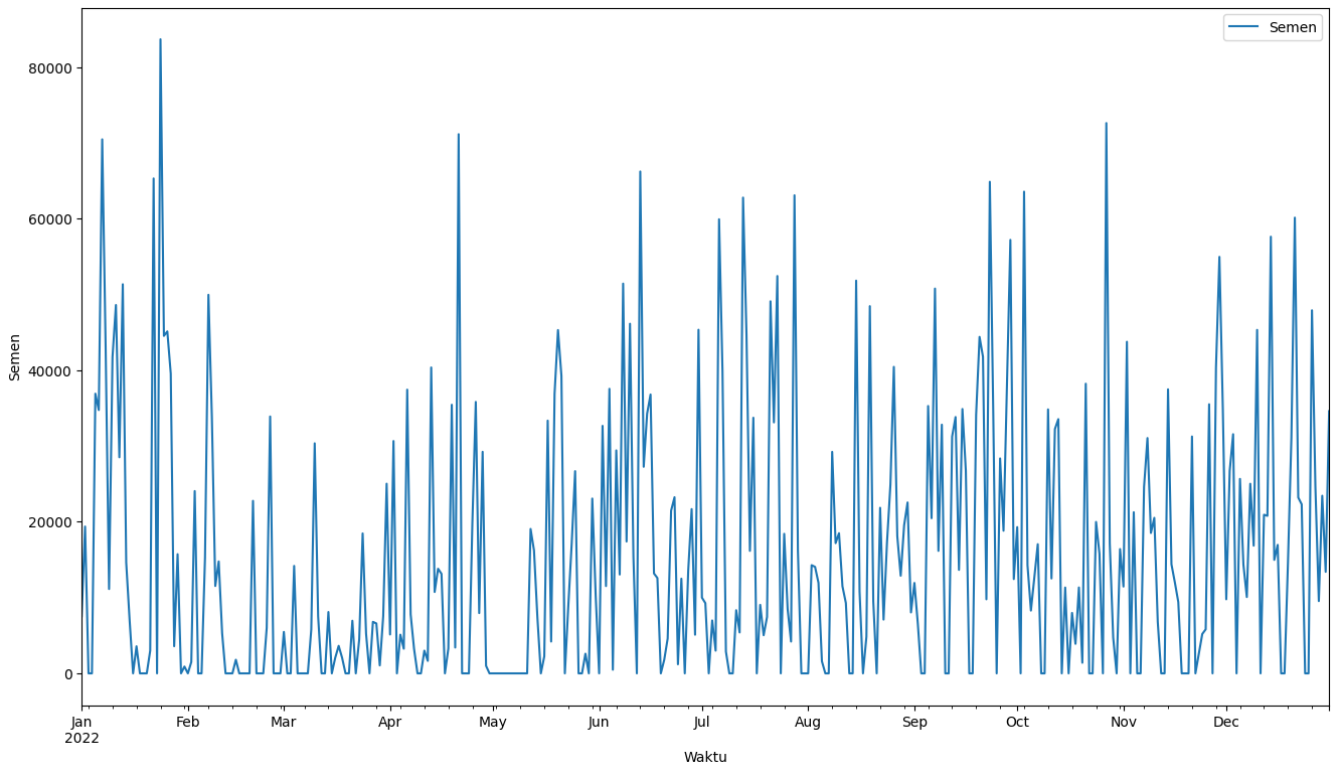
Penelitian dimulai dari mengidentifikasi gejala masalah empiris, yaitu proses pembuatan beton ready mix yang sering terlambat karena banyaknya permintaan dengan jumlah yang tidak menentu sehingga persediaan semen lebih cepat habis. Setelah itu dilakukan studi literatur terkait penelitian–penelitian terkait metode peramalan dan peramalan kebutuhan semen. Studi literatur ini bertujuan untuk memahami berbagai pendekatan yang telah digunakan dalam penelitian sebelumnya serta menemukan metode yang paling sesuai untuk diterapkan dalam konteks penelitian ini. Selanjutnya dilakukan input data dan rekayasa data dimana variabel waktu dibagi menjadi beberapa *feature*, yaitu *dayofweek*, *quarter*, *month*, *dayofyear*, *dayofmonth*, dan *weekofyear*. Pemilihan *feature* ini dilakukan untuk menangkap pola musiman dan tren dalam data yang dapat memengaruhi prediksi kebutuhan semen. Pengolahan data menggunakan Google Colab dengan bahasa pemrograman Python, karena platform ini menawarkan fleksibilitas dalam komputasi berbasis *cloud* dan mendukung berbagai pustaka *machine learning* yang diperlukan dalam penelitian ini.

Langkah selanjutnya adalah melakukan data scaling dengan menggunakan MinMaxScaler. Data scaling ini dilakukan karena tiap *feature* memiliki nilai yang berbeda-beda, dari jumlah satuan hingga puluhan ribu, sehingga apabila tidak dilakukan *scaling* maka hasil yang diperoleh kurang optimal. Selain itu, *scaling* juga membantu algoritma *machine learning* bekerja lebih efisien dengan mengurangi dampak perbedaan skala antar variabel. Setelah *scaling* dilakukan, selanjutnya membagi data menjadi data train dan data test untuk uji *cross-validation*. Data untuk uji *cross-validation* akan dilakukan mulai bulan 8 sampai bulan 11, atau selama 4 bulan, sedangkan bulan 12 digunakan untuk data test-nya. Uji *cross-validation* dilakukan untuk mengetahui hasil peramalan berdasarkan pembagian bulan, sehingga kondisi *overfitting* dapat diantisipasi. Dengan cara ini, model dapat diuji pada berbagai subset data guna memastikan bahwa performanya tetap konsisten dalam berbagai kondisi waktu dan permintaan.

Selanjutnya penggunaan beberapa algoritma machine learning untuk peramalan, antara lain *Support Vector Regression*, *Linear Regression*, *Random Forest Regressor*, *Decision Tree Regressor*, *Ridge*, *Lasso*, *Elastic Net*, *Extreme Gradient Boosting* (XGB), dan XGBRF. Seluruh algoritma tersebut digunakan untuk mengolah uji *cross-validation* dan data test. Pemilihan algoritma yang beragam bertujuan untuk mengeksplorasi metode mana yang paling sesuai dengan karakteristik data yang dimiliki, karena setiap algoritma memiliki keunggulan dan keterbatasan masing-masing. Setelah pengujian selesai dilakukan, selanjutnya dilakukan perbandingan antara masing-masing algoritma, sehingga bisa diperoleh algoritma dengan nilai error terkecil. Untuk mengukur nilai error hasil peramalan, digunakan beberapa metrik, yaitu *root mean squared error* (RMSE), *mean absolute error* (MAE), dan *mean absolute percentage error* (MAPE). Dengan kombinasi metrik ini, penelitian dapat mengevaluasi keakuratan model secara lebih komprehensif, baik dalam hal kesalahan absolut, persentase kesalahan, maupun sensitivitas terhadap nilai ekstrem dalam data. Terakhir adalah memberikan kesimpulan dari hasil penelitian yang telah dilakukan, serta mengevaluasi kemungkinan perbaikan model untuk pengembangan lebih lanjut, seperti eksplorasi data tambahan, tuning parameter algoritma, atau penggunaan teknik ensemble untuk meningkatkan akurasi peramalan.

HASIL DAN PEMBAHASAN

Pola data kebutuhan semen pada PT. X selama setahun dapat dilihat pada Gambar 2 dibawah. Pada pola tersebut, terlihat jelas bahwa kebutuhan maksimal dalam sehari bisa mencapai lebih dari 80000 kg semen, sedangkan juga terdapat hari dimana kebutuhan semen 0 kg, yang artinya tidak terdapat aktivitas pembuatan *ready mix*.



Gambar 2. Grafik pola kebutuhan semen harian

Data yang telah dimasukkan selanjutnya diolah atau direkayasa menjadi beberapa feature waktu, dari sebelumnya hanya memiliki satu feature saja, dengan format DDMMYYYY. Rekayasa data ini sangat penting dilakukan apabila menggunakan algoritma machine learning karena algoritma tersebut tidak dapat membaca feature waktu sebagai suatu kesatuan, melainkan data waktu tersebut harus dibagi-bagi terlebih dahulu. Pembagian ini dilakukan agar model dapat menangkap pola tren dan musiman yang mungkin tersembunyi dalam data. Misalnya, beberapa pola permintaan semen mungkin lebih sering terjadi pada hari-hari tertentu dalam seminggu atau bulan-bulan tertentu dalam setahun. Dengan membagi feature waktu menjadi *dayofweek*, *quarter*, *month*, *dayofyear*, *dayofmonth*, dan *weekofyear*, model dapat lebih mudah mengenali pola tersebut dan membuat prediksi yang lebih akurat. Selain itu, rekayasa feature ini juga membantu dalam mengurangi dimensi data yang tidak relevan, sehingga meningkatkan efisiensi dan performa algoritma yang digunakan. Akibatnya, feature–feature waktu yang tidak memiliki pengaruh kuat dapat dikeluarkan, dan hanya fokus saja pada feature dengan pengaruh cukup kuat, sehingga hasil prediksi yang dihasilkan lebih optimal. Hasil rekayasa feature dapat dilihat pada Tabel 1, yang menunjukkan bagaimana data waktu yang semula berbentuk tunggal telah dikonversi menjadi beberapa feature yang lebih informatif untuk proses analisis lebih lanjut.

Tabel 1. Potongan data dengan *feature* yang telah direkayasa

Tanggal	dayofweek	quarter	month	dayofyear	dayofmonth	weekofyear	Semen
01/01/2022 00:00	5	1	1	1	1	52	7540
02/01/2022 00:00	6	1	1	2	2	52	19382
03/01/2022 00:00	0	1	1	3	3	1	0
04/01/2022 00:00	1	1	1	4	4	1	0
05/01/2022 00:00	2	1	1	5	5	1	36920
06/01/2022 00:00	3	1	1	6	6	1	34764
07/01/2022 00:00	4	1	1	7	7	1	70492
08/01/2022 00:00	5	1	1	8	8	1	43089
09/01/2022 00:00	6	1	1	9	9	1	11134
10/01/2022 00:00	0	1	1	10	10	2	41764

Pada data *scaling* di Tabel 2, tidak ditemukan adanya *feature* tahun, karena proses *scaling* hanya dapat dilakukan apabila dalam satu *feature* terdapat data yang berbeda, dalam hal ini tahun yang berbeda, misal 2021, 2022, dan 2023. Namun pada data yang diperoleh hanya terdapat satu tahun saja yaitu 2022, sehingga *feature* tahun dapat dihapus. *Feature* semen juga wajib dilakukan *scaling* untuk menghindari perbedaan nilai antara *feature* waktu dan semen itu sendiri. Setelah dilakukan *scaling*, selanjutnya dilakukan pemisahan dimana *feature* waktu menjadi variabel X, dan *feature* semen menjadi variabel y. Masing – masing variabel juga kemudian dipisah lagi menjadi *data train* dan *data test*.

Tabel 2. *Feature* yang telah di-*scaling*

Tanggal	dayofweek	quarter	month	dayofyear	dayofmonth	weekofyear	Semen
01/01/2022 00:00	0.833	0.000	0.000	0.000	0.000	1.000	0.090
02/01/2022 00:00	1.000	0.000	0.000	0.003	0.033	1.000	0.232
03/01/2022 00:00	0.000	0.000	0.000	0.005	0.067	0.000	0.000
04/01/2022 00:00	0.167	0.000	0.000	0.008	0.100	0.000	0.000
05/01/2022 00:00	0.333	0.000	0.000	0.011	0.133	0.000	0.441
06/01/2022 00:00	0.500	0.000	0.000	0.014	0.167	0.000	0.415
07/01/2022 00:00	0.667	0.000	0.000	0.016	0.200	0.000	0.842
08/01/2022 00:00	0.833	0.000	0.000	0.019	0.233	0.000	0.515
09/01/2022 00:00	1.000	0.000	0.000	0.022	0.267	0.000	0.133
10/01/2022 00:00	0.000	0.000	0.000	0.025	0.300	0.020	0.499

Proses selanjutnya dilakukan olah data untuk masing – masing pembagian uji *cross-validation* untuk masing – masing algoritma. Hasil pengolahan dapat dilihat pada Tabel 3. Nilai yang dihasilkan pada Tabel 3 merupakan nilai untuk setiap algoritma dan menggunakan dua macam metric pengukuran, yaitu RMSE dan MAE. Penggunaan dua macam metric ini dianggap lebih tepat untuk melihat seberapa besar penyimpangan yang terdapat pada hasil uji. Nilai yang diperoleh juga dalam bentuk *scaling*, dengan nilai antara 0 – 1, sehingga pembacaan tingkat kesalahan lebih mudah dilakukan. Kemudian dilakukan pelabelan dengan pemberian warna untuk mengetahui urutan algoritma dan metric dengan tiga nilai *error* terendah. Hal ini dilakukan agar lebih mudah dalam mengidentifikasi hasil pengujian yang diperoleh. Warna kuning digunakan untuk hasil pengujian terendah pertama, warna biru digunakan untuk hasil pengujian terendah kedua, dan warna merah digunakan untuk hasil pengujian terendah ketiga. Karena hasil yang diperoleh dari setiap pengujian tidak akan berubah walaupun pengujian dilakukan berulang kali, maka setiap pengujian cukup dilakukan sekali saja. Pada bulan 12 tidak lagi dilakukan pengujian *cross-validation* karena tidak terdapat lagi data pada bulan selanjutnya yang dapat dijadikan sebagai *data test*. Hasil pengujian algoritma yang memiliki nilai *error* terendah merupakan algoritma yang paling tepat untuk diterapkan dalam peramalan kebutuhan semen ini.

Tabel 3. Hasil *cross-validation* untuk tiap algoritma

Algoritma	CV Bulan	8		9		10		11	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
SVR		0.165877	0.120711	0.250275	0.200168	0.290817	0.259917	0.209018	0.154826
LR		0.164869	0.135347	0.218998	0.173430	0.231484	0.203743	0.182413	0.163139
RF		0.218461	0.170546	0.197261	0.164591	0.254388	0.200448	0.216265	0.163709
DT		0.354377	0.279064	0.247680	0.232973	0.472436	0.411807	0.316609	0.223073
Ridge		0.153861	0.114138	0.237641	0.180966	0.221224	0.190982	0.171343	0.152744
Lasso		0.163560	0.122104	0.246416	0.201625	0.218835	0.167681	0.193097	0.170272
EN		0.163193	0.122794	0.243201	0.200128	0.220805	0.173441	0.194273	0.172432
XGB		0.234164	0.187097	0.202285	0.175447	0.262948	0.212788	0.174750	0.136276
XGBRF		0.207993	0.169434	0.206960	0.171059	0.262662	0.221309	0.181175	0.150949

	Terendah 1
	Terendah 2
	Terendah 3

Selanjutnya dilakukan perbandingan hasil untuk tiap algoritma yang digunakan. Hasil yang diperoleh adalah algoritma Ridge memiliki hasil terbaik dengan tiga kali menempati urutan pertama, baik untuk metric RMSE maupun MAE, dan tiga kali pula menempati urutan ketiga, sehingga totalnya sebanyak enam kali. Kemudian diurutan kedua dengan total sebanyak empat kali adalah algoritma XGBRF, dimana algoritma ini menempati urutan kedua dan ketiga sebanyak masing – masing dua kali, walaupun tidak pernah menempati urutan pertama. Algoritma lainnya hanya mampu menghasilkan kurang dari empat kali, bahkan algoritma Decision Tree tidak mampu menghasilkan performa yang baik untuk data kebutuhan semen ini. Dengan demikian, data kebutuhan semen di PT. X paling tepat apabila diramalkan menggunakan algoritma Ridge karena berhasil memperoleh nilai *error* yang paling kecil jika dibandingkan dengan algoritma – algoritma *machine learning* lainnya.

Tabel 3. Hasil perbandingan tiap algoritma

	Terendah 1	Terendah 2	Terendah 3	Total
SVR	1	0	0	1
LR	0	0	3	3
RF	0	2	0	2
DT	0	0	0	0
Ridge	3	0	3	6
Lasso	2	0	1	3
EN	0	3	0	3
XGB	1	2	0	3
XGBRF	0	2	2	4

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh bahwa algoritma *Ridge* memiliki nilai kesalahan yang paling kecil diantara algoritma lainnya setelah dilakukan uji *cross-validation* dengan nilai RMSE masing – masing sebesar 0,153861, 0,171343, dan nilai MAE sebesar 0,114138. Dengan demikian, algoritma *Ridge* adalah metode yang paling tepat untuk digunakan dalam proses peramalan kebutuhan semen ini. Selain itu, untuk penelitian lanjutan dapat dipertimbangkan menggunakan algoritma *deep learning*, seperti *recurrent neural network*, *long short-term memory*, atau juga menggunakan optimasi *hyperparameter* pada algoritma – algoritma *machine learning* yang digunakan pada penelitian ini untuk memperoleh hasil yang lebih baik. Dapat juga dipertimbangkan untuk memasukkan *feature – feature* lainnya yang memiliki hubungan dengan peningkatan pemakaian semen pada PT. X, sehingga hasil yang diperoleh lebih baik lagi.

DAFTAR PUSTAKA

- [1] Informasi Statistika Konstruksi. Kementerian Pekerjaan Umum dan Perumahan Rakyat.2022 – 2023
- [2] Adha, Idharmahadi. "Pemanfaatan abu sekam padi sebagai pengganti semen pada metoda stabilisasi tanah semen." *Rekayasa: Jurnal Ilmiah Fakultas Teknik Universitas Lampung* 15.1 (2011): 33-40.
- [3] Wijaya, Benarsetya Putra. *Analisis Biaya Kontruksi Jalan Beton Dan Jalan Aspal Dengan Metode Binamarga Selama Umur Rencana 20 Tahun*. Diss. Universitas Narotama, 2024.
- [4] Pratama, Fariz Yudha. *Perbandingan Peramalan Penjualan Peralatan Kardiologi Menggunakan Metode Machine Learning, Dan Statistika Pada PT. XYZ*. Diss. Institut Teknologi Sepuluh Nopember, 2022.
- [5] Achmad, Andi Muhammad Fiqri. *Peramalan Pupuk Urea Menggunakan Algoritma Prophet*. Journal of Agro-Industrial Engineering. Politeknik ATI Makassar. 2022.
- [6] Purnama, Drajat Indra, and Oki Prasetya Hendarsin. "Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR)." *Jambura Journal of Mathematics* 2.2 (2020): 49-59.
- [7] Wirdhaningsih, KurniartiPutri. *Penerapan Data Mining Algoritma Decision Tree C5. 0 Untuk Peramalan Forex*. Diss. Universitas Brawijaya, 2013.

- [8] Maulana, Sandy Andika, Shabrina Husna Batubara, and Wahyu Kurnia Rahman. "Penerapan Metode Naïve Bayes dalam Peramalan Polusi Udara di Kota Jakarta." *Mutiara: Jurnal Penelitian dan Karya Ilmiah* 1.6 (2023): 347-362.