

## Machine learning untuk Prediksi Produksi Gula Nasional

Toniyah Jaelani<sup>a</sup>, Mohamad Yamin<sup>a</sup>, Cokorda Prapti Mahandari<sup>a</sup>

<sup>a</sup>Program Studi Teknik Mesin, Fakultas Teknologi Industri, Universitas Gunadarma, Jl. Margonda Raya 100, Kota Depok, Indonesia  
e-mail: toniyahjaelani@gmail.com

### Kata kunci:

Machine learning,  
Prediksi,  
Produksi  
Gula

### ABSTRAK

Gula merupakan salah satu bahan utama yang selalu dibutuhkan untuk membuat makanan dan minuman. Saat ini produksi gula belum mampu memenuhi kebutuhan gula nasional. Maka dari itu, pemerintah terus menambah jumlah pabrik dan meningkatkan produktivitas pabrik gula yang ada untuk memenuhi kebutuhan gula nasional. Di sisi lain, produksi gula juga berfluktuasi dimana hal ini dapat menyebabkan terjadinya kelangkaan gula jika tidak diprediksi secara tepat dan akurat. Oleh karena itu, pada penelitian ini dilakukan perbandingan prediksi produksi gula Indonesia dengan menggunakan metode *machine learning* yaitu *Long Short-Term Memory (LSTM)* dan juga dilakukan prediksi dengan menggunakan metode regresi linier. Penelitian ini dilakukan berdasarkan data sekunder yang bersumber dari hasil penelitian dan laporan dari instansi terkait. Data yang digunakan adalah data nasional berupa *time series* selama 52 tahun yaitu dari tahun 1968 sampai dengan tahun 2020. Hasil penelitian ini menunjukkan bahwa *error* pada metode regresi linier adalah 8%. Hasil prediksi dengan *machine learning* menunjukkan *error* yang lebih kecil dibandingkan dengan metode regresi linier. Metode LSTM menghasilkan *error train* sebesar 0,069% dan nilai *error* data pengujian sebesar 0,082%. Hasil peramalan dari regresi linier memiliki *trend* produksi yang meningkat, namun pada metode LSTM hasilnya mengalami *trend* penurunan.

### Keyword:

Machine learning,  
Forecasting, Sugar  
Production

### ABSTRACT

Sugar is one of the main ingredients that are always needed to make food or drinks. Currently, sugar production has not been able to meet the national sugar demand. Therefore, the government continues to increase the number of factories and increase the productivity of existing sugar factories to meet national sugar needs. On the other hand, sugar production also fluctuates which can lead to sugar shortages if it is not predicted precisely and accurately. Therefore, in this study, a comparison of predictions of Indonesian sugar production was carried out using the machine learning method, namely *Long Short-Term Memory (LSTM)* and predictions were also made using the linear regression method. This research was conducted based on secondary data sourced from research results and reports from related agencies. The data used is national data in the form of *time series* for 52 years, from 1968 to 2020. The results of this study indicate that the error in the linear regression method is 8%. Prediction results with machine learning show a smaller error than the linear regression method. The LSTM method produces a train data error of 0.069% and the test data error value of 0.082%. Forecasting results from linear regression have an increasing trend of production, but the LSTM method results in a decreasing trend.

## 1. PENDAHULUAN

Gula merupakan salah satu bahan baku yang tak bisa lepas dari kebutuhan manusia sehari-hari, mulai dari gula kristal putih (GKP) yang dikonsumsi oleh masyarakat maupun gula rafinasi yang banyak dibutuhkan oleh industri makanan dan minuman. Kebutuhan gula juga mengalami fluktuasi dari tahun ke tahun seperti ketika mendekati momen-momen besar di Indonesia. Pada tahun 2021, kebutuhan gula konsumsi sebesar 646,9 ribu ton dan jika ditambahkan dengan kebutuhan gula rafinasi mencapai angka 2,1 juta ton sampai Bulan Mei dan akan bertambah menjadi 3,9 juta ton kebutuhan gula *raw sugar* impor [1].

Produksi gula dunia sekitar 70% berasal dari tanaman tebu, namun produksi gula tidak selalu menghasilkan jumlah yang sama dalam tiap tahunnya dikarenakan berbagai macam faktor [2]–[4]. Seperti Andi A. Dkk. telah melakukan penelitian dan menemukan berbagai macam faktor yang menyebabkan penurunan produksi gula ini yang meliputi, menurunnya area lahan tebu dan produktivitas tebu, kebun tebu tidak efisien, produksi tebu sedikit, kurangnya kualitas tebu, dan tidak efisiennya produksi tebu pada perusahaan gula [5].

Berdasarkan hal tersebut dibutuhkan proses peramalan produksi gula demi mendapatkan jumlah yang tepat dalam pemenuhan permintaan gula nasional. Peramalan adalah suatu aktivitas untuk memprediksi peristiwa masa depan dengan menggunakan data historis [6]. Banyak metode dalam ilmu statistik untuk digunakan sebagai alat peramalan data *time series* seperti *Smoothing*, Box-Jenkins (ARIMA), ekonometrik, regresi, dan fungsi transfer [7], [8]. Selain itu dalam *machine learning* terdapat juga beberapa metode untuk memprediksi seperti *Decision Tree Regressor*, *Random Forest Regressor*, *Artificial Neural Network* (ANN), *Recurrent Neural Network* (RNN), dan *Support Vector Machine* (SVM) [9].

Produksi gula mengalami fluktuasi dalam setiap tahunnya. Fluktuasi inilah yang bisa digunakan sebagai sinyal dalam menyusun rencana guna memenuhi kebutuhan gula Indonesia [10]. Dalam menyusun rencana tersebut dibutuhkan data historis untuk mengetahui apakah fluktuasi ini terjadi secara trend, seasonal ataupun *cyclic*. *Long Short-Term Memory* (LSTM) adalah salah satu metode RNN terbaik. Pada metode ini, RNN dimodifikasi untuk ditambahkan *memory cell* yang dapat menyimpan informasi untuk jangka waktu tertentu. *Recurrent Neural Network* (RNN) merupakan jenis arsitektur jaringan saraf tiruan yang berfungsi untuk proses memanggil berulang-ulang data input yang berupa data sekuensial. Data sekuensial memiliki karakteristik dimana sampel diproses dengan urutan waktu dengan suatu sampel memiliki keterkaitan dengan sampel lainnya. LSTM adalah salah satu algoritma RNN dimana terdapat *memory cell* untuk menyimpan informasi dalam jangka waktu yang lama.

Hasil dari suatu prediksi merupakan nilai kuantitatif terhadap subjek tertentu untuk jangka waktu tertentu. Prediksi selayaknya hanya sebagai bahan masukan terhadap suatu perencanaan. Kendati demikian, hasil dari prediksi dapat berbeda dari rencana yang diperkirakan. Pada metode peramalan terdapat 3 jenis metode yakni metode *time series* dimana teknik statistika yang menggunakan data historis sebagai dasar dari *forecasting*, metode regresi yang menghubungkan antara nilai produksi dan penyebabnya, dan yang ketiga adalah metode kualitatif dimana proses prediksi dilakukan dengan penilaian ahli, manajemen dan pendapat untuk memperkirakan hasil dari prediksi tersebut [11].

Maka dari itu, dalam penelitian ini dilakukan komparasi metode *machine learning* yakni RNN dengan algoritma LSTM dan metode regresi linier non *machine learning* prediksi produksi gula menggunakan indikator besaran *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE), sehingga dapat diketahui nilai keakuratan dari kedua metode tersebut.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini dihasilkan dari berbagai sumber didapat dari buku tahunan instansi terkait serta jurnal-jurnal terdahulu. Setelah itu, data dihimpun sesuai dengan variabelnya seperti tahun, produksi tebu, produksi gula, konsumsi gula, dan jumlah penduduk.

### 2.2 Preprocessing Data

Penggunaan LSTM pada penelitian ini dilakukan untuk komparasi hasil dari model prediksi regresi linier sederhana dan *machine learning*. Model yang dilakukan dihasilkan dari pengolahan dataset yang kemudian dataset tersebut akan dilakukan pengujian yakni *training* dan *testing* data.

Jumlah data yang digunakan terdiri dari 5 baris yaitu tahun, konsumsi gula, produksi tebu, produksi gula, serta jumlah penduduk dan 53 kolom yang berupa runtun waktu dari tahun 1968 sampai 2020. Berdasarkan jumlahnya, data tersebut dibagi dalam data *training* sebanyak 34 data dan data *testing* sebanyak 19 data.

### 2.3 Pemodelan LSTM

Pemodelan dalam penelitian ini menggunakan beberapa parameter yakni banyaknya *time step*, *epoch*, *batch size* serta *verbose*. Untuk mengetahui pemodelan tergolong baik atau tidak, peneliti menggunakan indikator *error* MSE dan MAE. Besaran nilai LSTM yang digunakan yakni 500. Hal ini dilakukan karena besaran *error* yang dihasilkan oleh LSTM 500 lebih baik jika dibandingkan dengan besaran lainnya.

### 2.4 Prediksi Train and Test Data

Pada prediksi data *train* dan data *test* dilakukan menggunakan *timestep*. *Timestep* adalah banyaknya data yang digunakan sebagai dasar dalam pemodelan prediksi. Pengujian terhadap data *train* dan data *test* dilakukan dengan melihat nilai MSE pada model prediksi.

### 2.5 Prediksi Nilai Masa Depan

Setelah didapatkan prediksi terhadap data *train* dan data *test*, selanjutnya adalah melakukan prediksi untuk 5 tahun ke depan terhadap produksi gula.

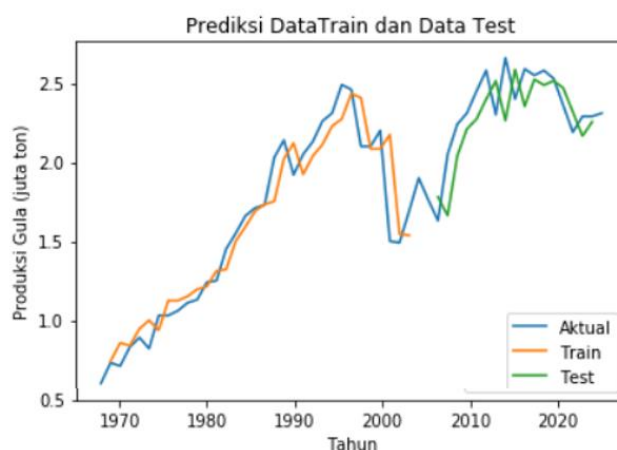
## 3. HASIL DAN PEMBAHASAN

Produksi gula dapat dipengaruhi oleh beberapa faktor di antaranya produksi tebu, konsumsi gula, dan jumlah penduduk. Data yang telah dilakukan pemodelan dalam dan di input kedalam algoritma LSTM dan RL menghasilkan nilai MSE pada data *train* dan data *test* yang ditunjukkan pada tabel di bawah ini.

Tabel 1. Nilai MSE dan MAE pada data *test* dan data *train* LSTM

Data	MSE	MAE	%Error
<i>Train</i>	1673048.13	1262.37	0.069%
<i>Test</i>	2297540.95	1511.18	0.082%

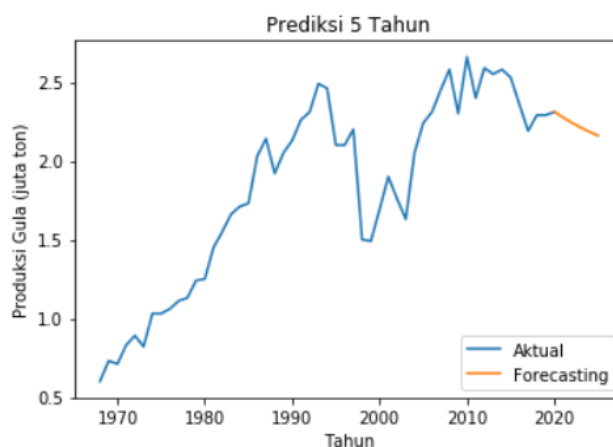
Dapat dilihat berdasarkan tabel 1 bahwa besaran *error* pada pemodelan data *train* sebesar 1673048,13 dan pada data *test* sebesar 2297540,95. Pada nilai MAE, data *train* menghasilkan *error* 1262,37 dan data *test* 1511,18. Persentase *error* pada data *train* yakni 0.069% dan pada data *test* 0.082%. Semakin kecil nilai *error* maka model prediksi semakin baik dan didapatkan nilai *error* pada proses pembuatan *forecasting* ini menandakan bahwa hasil yang akan didapatkan memiliki keakuratan data yang cukup baik.



Gambar 1. Hasil prediksi data train dan data test terhadap data valid

Pengujian data *train* dan data *test* dibandingkan dengan data actual, terlihat pada gambar 1, pemodelan data *train* dan data *test* hampir menyerupai data actual dimana pemodelan data *train* dan data *test* dilakukan untuk mengetahui nilai *error* pada algoritma LSTM yang digunakan dalam proses mendapatkan prediksi model prediksi cukup baik.

Berdasarkan data *train* dan data *test* yang sudah disimulasikan dalam algoritma LSTM, maka didapatkan hasil *forecasting* produksi gula dalam 5 tahun ke depan. Hasil tersebut dapat dilihat pada tabel di bawah ini.



Gambar 2. Hasil prediksi LSTM *machine learning*

Pola data runtun waktu yang digunakan menghasilkan hasil *forecasting* seperti yang terlihat pada grafik di atas, terlihat terjadi tren penurunan produksi gula selama 5 tahun ke depan. Perlu diketahui bahwa data yang digunakan dalam pemodelan akan mempengaruhi hasil prediksi dimana semakin banyak dan lengkap data historis yang digunakan, akan membuat pemodelan data lebih baik sehingga hasil prediksi memiliki nilai error dan kepresisian yang baik.

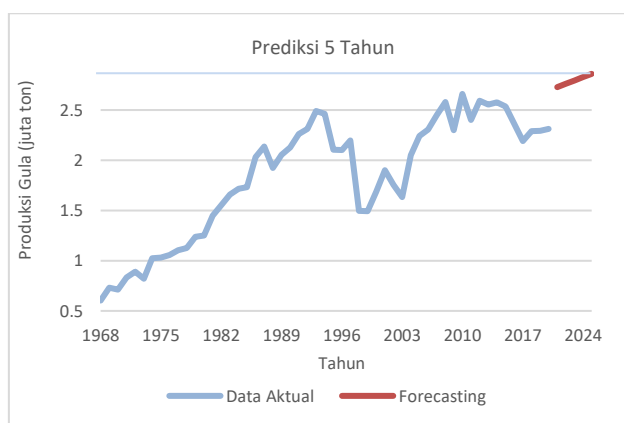
Selain itu dalam penelitian ini dilakukan metode regresi linier tanpa menggunakan *machine learning*. Besaran *error* yang terjadi pada pemodelan data *forecasting* menggunakan regresi linier *non machine learning* dapat dilihat pada tabel di bawah ini:

Tabel 2. Nilai MSE dan MAE pada Regresi Linier (RL) *non machine learning*

Data	MSE	MAE	%Error
RL	3.96E+10	151521.28	8%

Nilai MSE dan MAE pada metode regresi linier *non machine learning* lebih besar, dimana nilai MAE sebesar 3.96E+10 dan MAE sebesar 151521.28 dengan persentase error sebesar 8%. Data error yang terjadi pada pemodelan data memiliki nilai yang lebih besar jika dibandingkan dengan metode *machine learning* LSTM. Hal ini dapat terjadi karena pemodelan pada regresi linier *non machine learning* tidak memiliki keakuratan yang baik dan pemodelan yang tidak berulang seperti yang ada pada algoritma LSTM.

Berdasarkan model regresi linier tersebut dihasilkan *forecasting* seperti yang terlihat pada gambar 3 di bawah ini.



Gambar 3. Hasil prediksi RL *non machine learning*

Grafik di atas memperlihatkan hasil prediksi mengalami tren kenaikan produksi. Hal ini bertolak belakang dengan metode *machine learning* dimana hasil *forecasting* mengalami tren penurunan.

Pada tabel 3 dapat dilihat perbandingan hasil dari prediksi produksi gula menggunakan metode *machine learning* LSTM dan menggunakan regresi linier *non machine learning*.

Tabel 3. Perbandingan hasil prediksi LSTM dan RL *non machine learning*

Tahun	LSTM (ton)	RL (ton)
2021	2275647	2727900
2022	2242033	2760883
2023	2211829	2793866
2024	2184684	2826850
2025	2160286	2859833

Berdasarkan tabel perbandingan di atas, hasil dari LSTM mengalami penurunan pada tahun 2021. Hasil prediksi menunjukkan produksi gula sebesar 2,27 juta ton, kemudian penurunan berlanjut pada tahun 2022 dimana produksi gula sebesar 2,24 juta ton, pada tahun 2023 produksi gula menurun menjadi 2,21 juta ton, tahun 2024 penurunan gula terus berlanjut menjadi 2,18 juta ton, dan pada 2025 produksi gula sebesar 2,16 juta ton. Akan tetapi, pada metode regresi linier, hasil *forecasting* mengalami kenaikan. Hal bertolak belakang terjadi pada metode regresi linear dimana terjadi tren kenaikan produksi. Pada tahun 2021, produksi gula menghasilkan *forecasting* sebesar 2,73 juta ton, pada tahun 2022 berangsur naik sebesar 3 ribu ton menjadi 2,76 juta ton, dan pada tahun 2023 produksi gula naik sebesar 110 ribu ton pertahun menjadi 2,79 juta ton. Pada tahun 2024, produksi gula naik kembali menjadi 2,87 juta ton, dan pada tahun 2025 naik menjadi 2,86 juta ton pertahun. Ini bisa diakibatkan karena pemodelan yang dilakukan pada sistem yang cukup berbeda.

#### 4. KESIMPULAN

*Forecasting* sangat membantu dalam memperkirakan langkah seperti apa yang dibutuhkan dalam beberapa tahun kedepan. *Forecasting* terhadap produksi gula dari dua model prediksi yang telah dilakukan yaitu *machine learning* dengan LSTM dan *non machine learning* yang menggunakan regresi linier memiliki tren yang berbeda dimana *error* yang dihasilkan dari LSTM lebih kecil dibandingkan dengan metode regresi linier yakni *error* pada metode regresi linier adalah 8%. Sedangkan dalam metode LSTM menghasilkan *error data train* sebesar 0,069% dan nilai *error data test* sebesar 0,082%.

Pada model LSTM, tren produksi gula mengalami penurunan setiap tahunnya mulai dari 2,27 Juta ton pada tahun 2021 sampai dengan 2,16 juta ton pada tahun 2025. Sedangkan pada model regresi linier Produksi Gula mengalami kenaikan tren dari 2,72 juta ton pertahun sampai 2,85 juta ton pertahun pada 2025. Penelitian ini dapat dijadikan bahan acuan dalam membuat prediksi yang lebih mendalam seperti menambahkan berbagai macam variabel baru seperti tumbuhnya jumlah pabrik gula serta korelasinya terhadap hasil produksi gula.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa yang telah memudahkan jalannya penelitian ini, Universitas Gunadarma yang telah mengikutsertakan pada penulisan ilmiah, kedua orang tua dan dosen pembimbing yang tiada hentinya memberikan arahan, masukan dan saran untuk menyelesaikan penelitian ini.

#### REFERENSI

- [1] B. P. Statistik, *Distribusi Perdagangan Komoditas Gula Pasir*. 2021.
- [2] M. Z. Hassan, M. Siraj-Ud-Doulah, and M. K. Hasan, "Forecasting The Production of Sugar Cane Based on Time series Models in Bangladesh," *Bull. Math. Stat. Res.*, vol. 7, no. 4, pp. 77–89, 2019, doi: 10.33329/bomsr.74.24.
- [3] A. K. Srivastava and M. K. Rai, "Sugarcane production: Impact of climate change and its mitigation," *Biodiversitas, J. Biol. Divers.*, vol. 13, no. 4, pp. 214–227, 2012, doi: 10.13057/biodiv/d130408.

- [4] D. Zhao and Y. R. Li, "Climate Change and *Sugarcane* Production: Potential Impact and Mitigation Strategies," *Int. J. Agron.*, vol. 2015, 2015, doi: 10.1155/2015/547386.
- [5] A. A. Sulaiman, Y. Sulaeman, N. Mustikasari, D. Nursyamsi, and A. M. Syakir, "Increasing *Sugar* Production in Indonesia through Land Suitability Analysis and *Sugar* Mill Restructuring," *MDPI*, vol. 8, no. 61, pp. 1–17, 2019.
- [6] A. Kantasa-ard, M. Nouiri, A. Bekrar, A. Ait el cad, and Y. Sallez, "Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand," *Int. J. Prod. Res.*, vol. 59, no. 24, pp. 7491–7515, 2021, doi: 10.1080/00207543.2020.1844332.
- [7] I. Harlianingtyas, A. Salim, D. Hartatie, and S. Supriyadi, "Forecasting *sugarcane* production in the Asembagus *sugar* factory," in *IOP Conference Series: Earth and Environmental Science*, 2020, vol. 411, no. 1, p. 12022, doi: 10.1088/1755-1315/411/1/012022.
- [8] F. N. Fauziah and A. Gunaryati, "Comparison forecasting with double exponential smoothing and artificial neural network to predict the price of *sugar*," *Int. J. Simul. Syst. Sci. Technol.*, vol. 18, no. 4, pp. 11–13, 2017, doi: 10.5013/IJSSST.a.18.04.13.
- [9] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," *Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015*, pp. 2823–2824, 2015, doi: 10.1109/BigData.2015.7364089.
- [10] M. A. Heryanto and E. R. Suryatmana, "Dinamika Agroindustri Gula Indonesia: Tinjauan Analisis Sistem," *Agricore J. Agribisnis dan Sos. Ekon. Pertan. Unpad*, vol. 5, no. 2, 2020.
- [11] J. Christy, A. P. Hintarsyah, and H. L. H. Spits Warnars, "Forecasting Sebagai Decision Support Systems Aplikasi dan Penerapannya Untuk Mendukung Proses Pengambilan Keputusan.," *J. Sist. Komput.*, vol. 8, no. 1, pp. 19–27, 2018.