

Submitted : 25 November 2023

Revised : 17 April 2024

Accepted : 21 May 2024

DEEP LEARNING DENGAN METODE LSTM (LONG SHORT-TERM MEMORY) UNTUK PEMODELAN KELAKUAN DINAMIK SISTEM DISTILASI MENGGUNAKAN REAL OPERATING DATA

William Kurniawan*, Budi Husodo Bisowarno

Program Studi Teknik Kimia, Fakultas Teknologi Industri, Universitas Katolik Parahyangan,
Bandung, 40141, Indonesia

*Email: 8142101003@unpar.ac.id

Abstrak

Kemurnian produk tidak dapat diukur secara *on-line* dan terdapat *time delay* dalam pengukuran kemurnian secara *offline* dan tidak kontinu dengan hasil analisis kemurnian dari laboratorium. Kontrol berbasis variabel inferensial merupakan alternatif untuk memprediksi dan mengendalikan kemurnian produk dengan variabel kondisi operasi yang tepat secara *real-time*. Pabrik memiliki data kondisi operasi (*input*) serta data kemurnian produk (*output*) berasal dari kondisi nyata pabrik dapat digunakan untuk pembangunan model *deep learning*. Penelitian ini dibagi menjadi tiga tahapan besar, yaitu pertama adalah tahapan persiapan data untuk menentukan variabel paling sensitif dalam memprediksi kemurnian produk menggunakan metode korelasi pearson yaitu temperatur *upper column* dan *process knowledge* yaitu *pressure top column*. Tahapan kedua adalah penyusunan algoritma *deep learning* menggunakan metode LSTM (*long short-term memory*) menggunakan bantuan program python. *Optimizer* yang digunakan untuk melatih data adalah *optimizer* SGD (*stochastic gradient descent*). Data yang digunakan untuk melatih dan melakukan validasi model *deep learning* adalah data dari 2018 hingga semester pertama tahun 2019 dengan rasio training set:test data = 70:30. Tahapan ketiga adalah melatih model dan validasi model, memprediksi hasil, serta melakukan evaluasi model terhadap data pabrik terhadap set data semester kedua tahun 2019. Hasil dari pelatihan adalah model LSTM dengan nilai RMSE (*root mean squared error*) sebesar 1,4. Nilai validasi untuk model ini adalah RMSE sebesar 0,6389. Model ini dapat memprediksi set data semester kedua 2019 dengan nilai $R^2 = 0,74$ dan RMSE = 0,56.

Kata Kunci: *Deep learning*; LSTM; Pemodelan proses; Sistem distilasi

Abstract

Industrial plants experience difficulties maintaining product purity from the distillation system. The purity of the products cannot be measured online, and there is a time delay in offline and non-continuous purity measurements with laboratory purity analysis results. Control based on inferential variables is an alternative to predicting and controlling product purity with operational condition variables in real-time. The plant has operational condition data (input) and product purity data (output) from actual plant conditions that can be used for developing deep-learning models. The first stage is determining the most sensitive variables in predicting product purity using Pearson correlation methods, i.e., Upper Column Temperature, and process knowledge, i.e., Top Column Pressure. The second stage involves developing a deep learning algorithm using Python programming, using LSTM (Long-Short Term Memory) methods. The optimizer used while training is SGD (Stochastic Gradient Descent). The data used for training and validating the deep learning model are from 2018 to the first half of 2019 with a training set:test data ratio of 70:30. The third stage is training and validating the model, predicting results, and evaluating the model against plant data for the second semester of 2019. The best training result is an LSTM model with an RMSE value of 1.4. The validation value for this model is an RMSE of 0.6389. This model can predict the data set for the second semester of 2019 with an R^2 value of 0.74 and an RMSE of 0.56.

Keywords: *Deep learning*; Distillation system; LSTM; Process modelling

1. PENDAHULUAN

Salah satu proses penting dalam sebuah pabrik kimia adalah proses distilasi. Distilasi merupakan proses pemurnian senyawa cair-cair yang memerlukan energi yang besar, yaitu sekitar 40% total energi dalam proses pabrik kimia (Oda et al., 2012). Energi tersebut digunakan untuk mengubah fase dengan penguapan dan kondensasi mendapatkan tahapan kesetimbangan uap-cair sehingga diperoleh produk murni.

Pabrik mengalami kesulitan dalam mendapatkan serta mempertahankan kemurnian produk A yang diproses menggunakan sistem distilasi. Kemurnian produk dari suatu sistem distilasi sangat dipengaruhi oleh kondisi operasi dari kolom, yaitu variabel tekanan, temperatur, serta laju alir pada kolom distilasi. Sistem distilasi pada pabrik ini merupakan sistem dinamik yang mana variabel kondisi operasi berubah sepanjang waktu. Sistem proses distilasi yang dikaji dapat dilihat pada Gambar 1 dengan variabel proses disajikan pada Tabel 1.

Komposisi produk pada kolom tidak dapat diukur secara *online/real time*. Kemurnian produk dianalisis di laboratorium dengan metode *gas chromatography (GC)*. Kekurangan dari metode analisis laboratorium adalah timbul suatu *time delay*. *Time delay* yang dimaksud mencakup waktu untuk pengambilan sampel yang kemudian diantar ke laboratorium, kemudian waktu untuk menganalisis sampel, dan waktu untuk melaporkan kembali kepada pihak produksi pabrik. Tentunya kekurangan ini menjadi kesulitan dalam pengoperasian pabrik.

Tabel 1. Variabel pada sistem proses

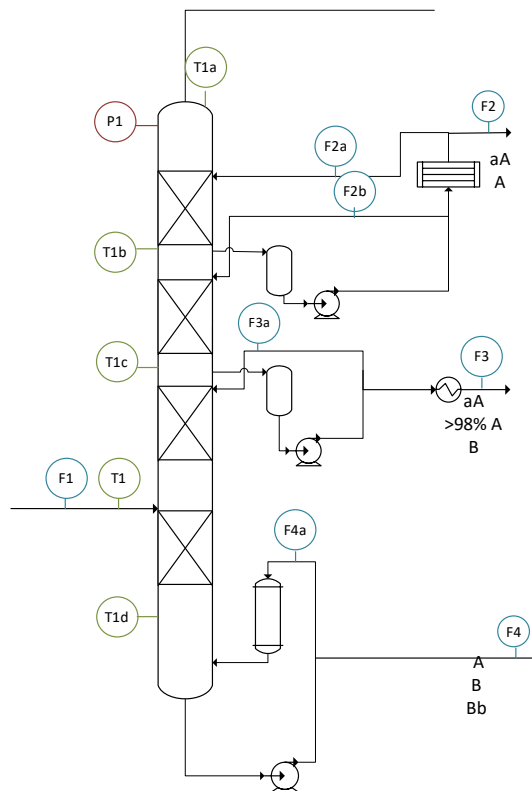
Simbol	Variabel
T1	Temp. Feed Column 1
T1A	Temp. Top Column 1
T1B	Temp. Upper Column 1
T1B	Mid Temperature Column 1
T1C	Temp. Bottom Column 1
P1	Pressure Top Column 1
F1	Flow Feed Column 1
F2	Flow Product Overhead Column 1
F2A	F Reflux Overhead Column 1
F2B	F Pumpharound Overhead Column 1
F3	F Product Mid Column 1
F3A	F Pumpharound Mid Column 1
F4	F Product Bottom Column 1
F4A	Flow Pumpharound Bottom Column 1
W1	Product A composition

Untuk dapat mengukur kemurnian secara *real time/online*, diperlukan peralatan instrumentasi tambahan yang harus diinstalasi pada sistem kolom distilasi. Penambahan peralatan ini tentu memerlukan biaya kapital serta biaya perawatan selama beroperasi. Oleh karena itu, pabrik memutuskan untuk tetap menggunakan analisis laboratorium untuk menganalisis kemurnian produk.

Penerapan berdasarkan prinsip hukum konservasi tidak bisa selalu dilakukan untuk memodelkan kondisi dinamik proses di pabrik. Contoh di pabrik ini adalah faktor ketidakakuratan dari pembacaan instrumentasi

seperti temperatur, tekanan, dan laju alir. Apalagi data yang digunakan bersifat dinamik dan hanya berlaku pada sistem proses tertentu. Selain itu, terdapat hubungan non-linier kompleks antar variabel dan bersifat dinamis akan sulit dimodelkan menggunakan prinsip konservatif (Park et al., 2022). Kelemahan dari *traditional learning* adalah pengguna harus memiliki pemahaman dari persoalan (Karthikeyan & Priyakumar, 2022) Sebagai alternatif, dengan memiliki data kondisi operasi (*input*) serta data kemurnian produk (*output*) berasal dari kondisi nyata pabrik dapat digunakan untuk pembangunan model proses. Paradigma penggunaan komputer sudah beralih dari *traditional programming* menjadi *machine learning* (Mowbray et al., 2022). Berdasarkan data pabrik untuk periode tahun 2019, metode *machine learning deep learning* akan diterapkan untuk melatih dan memvalidasi model proses dengan menggunakan set data tersebut, sehingga model proses dapat memprediksi kondisi operasi yang menghasilkan kemurnian produk.

Kontrol berbasis variabel inferensial merupakan alternatif untuk memprediksi kemurnian produk dengan variabel kondisi operasi temperatur (Venkateswarlu & Avantika, 2001). Apabila *disturbance* secara langsung mempengaruhi komposisi, maka kontrol secara inferensial menjadi permasalahan, maka *machine learning* merupakan metode paling tepat dalam mempelajari data-data non-linier.



Gambar 1. Sistem proses

Osuolale & Zhang, (2016) melakukan penelitian menggunakan ANN untuk mendapatkan kondisi operasi distilasi dengan target penggunaan eksergi secara maksimal. Peneliti lain menerapkan *deep*

learning untuk sistem 2,3-butadienol (Choi et al., 2023). Kwon et al. (2021) melakukan studi untuk memodelkan *machine learning-based prediction model* untuk sistem proses distilasi 99% normal butana dari campuran butana. Metode *deep learning* yang digunakan adalah RNN (*recurrent neural network*), LSTM (*long short-term memory*), dan GRU (*gated recurrent unit*) dengan metode optimasi adam, adagrad, RMSprop, dan SGD (*stochastic gradient descent*). Metode RNN dipilih karena dapat menganalisis dan memprediksi data dengan *time-series*. Hasil dari penelitian tersebut membuktikan bahwa model *machine learning* dapat digunakan untuk memprediksi perilaku kolom distilasi melalui temperatur kolom. Kelebihan dari penggunaan *machine learning* ini adalah model ini menggunakan data aktual proses dalam pelatihan model yang mana sangat reliabel dan dapat diaplikasikan langsung pada proses. *Deep learning* merupakan salah satu metode yang dapat mencari hubungan input-output antara data non-linier kompleks. RNN memiliki keterbatasan yang pada data-data yang saling berhubungan secara temporer, sehingga model LSTM merupakan model yang menutupi kekurangan RNN (Shewalkar, 2018)

Pabrik ini memiliki data selama tahun 2018-2019, yang cocok digunakan untuk melatih model *deep learning*. Instrumentasi yang terpasang dilakukan kalibrasi ulang setiap tahun untuk memastikan data akurat. Tidak ada perubahan signifikan pada kolom distilasi, serta tidak ada perubahan SOP pabrik, sehingga data selama tahun 2018-2019 dapat digunakan untuk melatih model *deep learning* untuk memprediksi kemurnian produk berdasarkan variabel proses kondisi operasi.

Fokus penelitian adalah mengembangkan *data-driven model* untuk sistem kolom distilasi produk A. Terdapat 14 buah data variabel yang terdiri dari 8 buah data variabel laju alir dalam satuan kg/h, 1 data variabel tekanan dalam satuan mbar, 5 buah data variabel temperatur dalam satuan °C, serta data kemurnian produk A yang diperoleh. Pemilihan variabel dan output akan dilakukan berdasarkan dari *process knowledge* dan divalidasi dengan menghitung korelasi pearson, yang dalam sistem proses ini adalah *input* berupa temperatur dan tekanan sistem distilasi, serta *output* berupa komposisi produk. Data-data variabel instrumentasi tersebut diukur secara *real-time* dan direkam setiap 1 jam sekali dan disimpan di *drive*. Untuk data *output* diambil setiap 4 jam sekali dengan metode analisis *gas chromatography*.

Data tersebut perlu diolah untuk menentukan data mana yang paling sensitif terhadap pengaruh kemurnian dalam kolom. Metode *preprocessing* yang digunakan adalah menggunakan korelasi pearson dan *min-max scaling*. Data juga kemudian dibagi menjadi 3, yaitu *set data* untuk melatih model *deep learning*, melakukan validasi pelatihan model, dan untuk melakukan prediksi kondisi operasi.

2. METODE

Penelitian dilakukan dalam tiga tahapan besar, yaitu pertama adalah tahapan persiapan data, yang

terdiri dari pengumpulan data, pemilihan variabel, serta melakukan *min-max normalization*. Tahapan kedua adalah penyusunan algoritma dari model *machine learning* menggunakan perangkat lunak python dengan memanfaatkan *library python* seperti: numPy, pandas, keras, dan scikit-learn, seaborn, scipy.

Setelah data di-*preprocess*, data akan diolah menggunakan algoritma *machine learning* yaitu LSTM (*long short-term memory*). Selama masa *training model*, model optimasi yang digunakan adalah SGD (*Stochastic Gradient Descent*). Tahap ketiga adalah melakukan *training model* serta evaluasi validasi model menggunakan set data tahun 2018 hingga semester pertama 2019 dengan *training:test = 70:30*. Model dievaluasi menggunakan RMSE dan R². RMSE merupakan parameter evaluasi yang memberikan nilai yang lebih seimbang terhadap kecocokan model. RMSE lebih sensitif terhadap nilai *error* lebih kecil (Chai & Draxler, 2014). Model dikatakan sempurna bila nilai RMSE mendekati nol. R² merupakan sebuah parameter yang merepresentasi tingkat kemiripan antara hasil prediksi dengan nilai sesungguhnya (Lee et al., 2021)

Model kemudian digunakan untuk melakukan prediksi hasil model yang telah dilatih dan dievaluasi menggunakan set data dari semester kedua tahun 2019 dan kemudian dievaluasi kembali menggunakan parameter RMSE dan R². Seluruh eksekusi program *python* dilakukan dengan menggunakan server google colab yang diakses pada November 2023.

2.1 Pengumpulan Data

Data yang diambil adalah data dari sebuah Pabrik dengan periode bulan Januari-Desember 2019. Sistem proses yang dikaji merupakan satu buah kolom distilasi untuk mengolah mendapatkan produk A dengan kemurnian tinggi. Data diperoleh dari sistem instrumentasi yang terinstalasi pada pabrik.

Perolehan data input diambil dari sistem DCS yang secara otomatis dihasilkan pada setiap jam, sebanyak 14 buah variabel, yang terdiri dari 8 variabel *flow* dalam kg/h, 1 variabel tekanan dalam mbar, dan 5 variabel data temperatur dalam °C. Untuk data output tersedia setiap 4 jam sekali, yaitu pada 02:00, 06:00, 10:00, 14:00, 18:00, dan 22:00. Data output didapat dari analisis laboratorium dengan metode GC untuk mengetahui komposisi dari produk A. Adapun asumsi yang digunakan selama proses pengumpulan data adalah:

- Data yang diambil dari sistem DCS (input), diasumsikan akurat dan menggambarkan data apa adanya. Instrumentasi ini dikalibrasi secara rutin setiap tahun untuk memastikan keakuratan data.
- Tidak ada modifikasi terhadap peralatan produksi, instrumentasi, dan SOP pengoperasian produksi, sehingga data tahun 2018-2019 dapat digunakan untuk *training set* model *deep learning* serta digunakan untuk validasi dan evaluasi.
- Data output berupa kemurnian, diasumsikan akurat dan mengabaikan *human error* yang mungkin terjadi. Peralatan *gas chromatography* di laboratorium dikalibrasi secara rutin setiap tahun untuk memastikan keakuratan data.

- Waktu pengambilan sampel adalah setengah jam sebelum waktu yang tertera (Contoh: Pada sampel pukul 10.00, maka waktu pengambilan sampel adalah 09:30). Kemudian asumsi waktu tinggal untuk masing-masing sampel adalah 30 menit. Maka data variabel kondisi input pada pukul 09:00 akan dicocokkan dengan data pukul 10:00
- Data output memiliki kecenderungan uptrend dan downtrend pada rentang waktu tertentu, oleh karena itu teknik interpolasi linier digunakan untuk menghasilkan data dalam kurun waktu tersebut.

2.2 Persiapan Data

Tahapan-tahapan persiapan data adalah mengekstrak *dataset* dari pangkalan data serta mengkaji struktur *dataset* terhadap tujuan penerapan *machine learning* (Ge et al., 2017).

Tahap berikutnya adalah proses koleksi data. Pada tahapan ini data dari pangkalan data diambil. Pangkalan data merekam data secara *real-time* melalui instrumentasi sensor DCS yang terinstalasi untuk mengontrol proses. Data yang diperoleh harus diformat dalam format .csv dalam bentuk tabel untuk memudahkan pelatihan model *machine learning* dengan perangkat lunak python. Variabel proses tersebut kemudian diseleksi menggunakan algoritma korelasi pearson yang tersaji pada persamaan (1).

$$R = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2} \sqrt{\sum_i^N (y_i - \bar{y})^2}} \quad (1)$$

dimana R adalah koefisien pearson, x_i dan y_i adalah variabel data dan \bar{x} dan \bar{y} adalah rata-rata variabel data tersebut.

Ada beberapa variabel yang tidak mempunyai pengaruh signifikan terhadap kemurnian produk. Jika variabel ini digunakan untuk melatih model, maka kualitas akurasi model akan menurun dan meningkatkan beban komputasi selama melatih model *machine learning*. *Key variable* tersebut akan dicari dengan melakukan analisis korelasi pearson yaitu antara variabel input (laju alir, temperatur, dan tekanan kolom) terhadap kemurnian produk.

Skala yang digunakan pada masing-masing variabel berbeda. Untuk meningkatkan kinerja dari model, maka data harus di-*preprocess* yaitu data yang diolah harus diubah menjadi rentang 0-1 dengan cara normalisasi. Salah satu cara normalisasi adalah menggunakan persamaan *min-max normalization*. Persamaan *min-max normalization* adalah sebagai berikut:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Dengan x' adalah nilai ternormalisasi, x_{min} dan x_{max} adalah nilai variabel data x minimum serta x maksimum.

Berikut merupakan langkah-langkah dalam penyusunan algoritma deep learning

1. Mengubah skala set data dengan *min-max*
2. *Split set data* menjadi *training:test data* dengan rasio 70:30
3. Mengubah permasalahan *multivariate time series* menjadi *supervised learning*

Setelah algoritma *deep learning* disusun di program python, maka langkah selanjutnya adalah melatih model menggunakan *set data* tahun 2018 hingga semester pertama 2019.

Fungsi aktivasi yang digunakan dalam *deep learning* adalah fungsi linear, yang secara langsung proposional terhadap nilai *input*. (Sharma et al., 2020). Metode SGD digunakan untuk melakukan optimasi terhadap *loss* pelatihan data untuk mendapatkan parameter pada model *deep learning* (Netrapalli, 2019).

2.3 Metode Evaluasi

Model yang telah dilatih kemudian digunakan untuk prediksi. Nilai *output* yang diperoleh dievaluasi dengan dua metode, yaitu R^2 dan RMSE. Nilai R memiliki rentang nilai dari angka 0 hingga 1, yang mana semakin mendekati nilai 1 maka prediksi model akan semakin akurat. Nilai RMSE menghitung nilai galat dari data asli dan data prediksi, yang mana semakin kecil nilai RMSE, maka model semakin akurat. Berikut persamaan R^2 dan RMSE.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (3)$$

$$R^2 = \left(\frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2} \sqrt{\sum_i^N (y_i - \bar{y})^2}} \right)^2 \quad (4)$$

Metode evaluasi model ini digunakan untuk menentukan kinerja dari setiap model. Validasi model dilakukan untuk mengkonfirmasi keabsahan dari sebuah model (Esbensen & Geladi, 2010). Model dengan nilai R^2 tertinggi dan RMSE terendah akan digunakan untuk melakukan prediksi menggunakan *set data* pada akhir semester 2019. Hasil prediksi kemudian dievaluasi kembali dengan metode yang sama, yaitu R^2 dan RMSE.

3. HASIL DAN PEMBAHASAN

Variabel data yang diperoleh dari sistem proses (Gambar 1) yang dikaji adalah sebanyak 14 variabel proses data dan 1 buah variabel kemurnian yang disajikan pada Tabel 1. Variabel data tersebut kemudian dipilih menggunakan *process knowledge* dan korelasi pearson untuk menentukan variabel proses mana yang paling dominan dalam memprediksi kemurnian produk. Jumlah data dari masing-masing variabel kondisi proses adalah sebanyak 17379 buah data dan sebanyak 2470 jumlah data. Data ini berasal dari sistem proses tahun 2018 hingga 2019.

3.1 Pearson Correlation Coefficient

Pearson correlation coefficient menghitung hubungan linier dari dua buah variabel. Sebanyak 14 buah variabel proses akan dihitung koefisien korelasi pearson terhadap produk A. Hasil dari perhitungan korelasi koefisien pearson disajikan pada Tabel 2.

Tabel 2. Korelasi pearson

Variabel Proses	Pearson A
Temperature Feed Column 1	0,240
Temperature Top Column 1	0,208
Temperature Upper Column 1	0,305
Temperature Middle Column 1	0,247
Temperature Bottom Column 1	0,187
Press Top Column 1	-0,124
Flow Feed Column 1	0,279
Flow Product Overhead Column 1	0,091
Flow Reflux Overhead Column 1	0,173
Flow Pumparound Overhead Column 1	0,144
Flow Product Mid Column 1	0,273
Flow Pumparound Mid Column 1	0,179
Flow Pumparound Bottom Column 1	0,158
Flow Feed Column 2	0,248
A	1

Hasil dari rentang korelasi pearson ini menunjukkan bahwa setiap variabel kondisi proses tidak memiliki hubungan linier terhadap kemurnian produk. Hasil dari korelasi pearson adalah variabel proses *temperature upper column* memiliki korelasi sebesar 0,305. Dalam penelitian ini, variabel paling dominan dipilih berdasarkan *process knowledge* pada sistem distilasi dimana kemurnian produk ditentukan oleh tekanan dan temperatur, maka dari itu variabel yang dipilih adalah *pressure top column* dan *temperature upper column*. Teori ini juga didukung dari hasil perhitungan korelasi pearson tertinggi yaitu *temperature upper column*. Nilai korelasi pearson pada variabel *pressure top column* bernilai negatif yang berarti memiliki hubungan terbalik terhadap nilai kemurnian produk.

Variabel yang memiliki korelasi kedua tertinggi adalah variabel *feed column*, yaitu sebesar 0,279. Variabel ini bukan merupakan *manipulated variable* sehingga nilai tidak dapat diubah. *Feed column*

ditentukan berdasarkan kapasitas pabrik sesuai dengan permintaan produk. Karena pearson hanya merupakan parameter hubungan linier, maka pearson ini tidak cocok secara langsung diterapkan untuk sistem distilasi pabrik ini. Model *deep learning* seperti LSTM lebih cocok untuk merepresentasikan korelasi antar variabel sistem distilasi proses.

Deep learning dengan LSTM dilatih dengan menggunakan *set data* kondisi operasi proses dengan *objective function* merupakan nilai minimal dari *loss function* (selisih antara data prediksi dengan data aktual). Artinya pelatihan model *deep learning* ini sudah disesuaikan dengan variabel *upper temperature column* dan *top pressure column*.

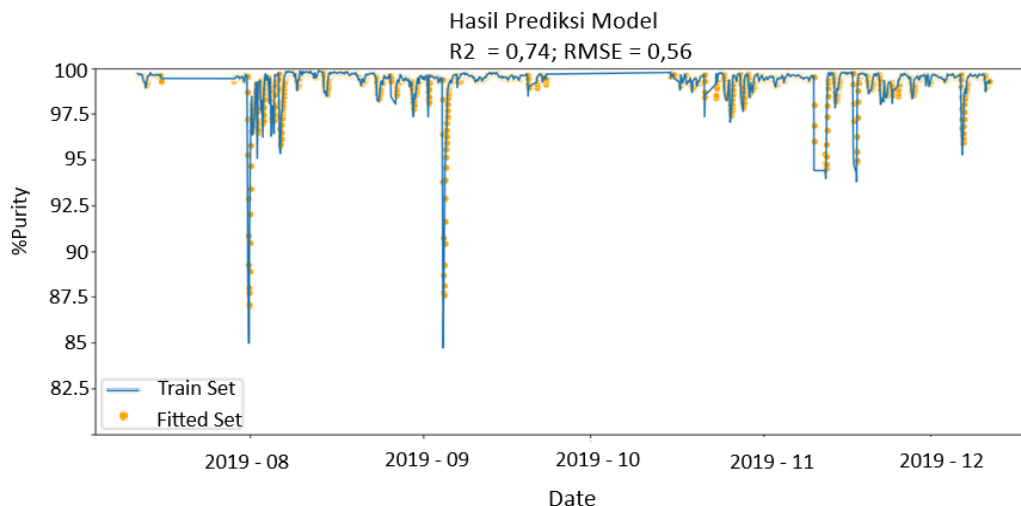
3.2 Data Cleaning & Formatting

Variabel kondisi proses memiliki jumlah sebanyak 17379 buah data dan sebanyak 9650 jumlah data kemurnian produk. Terdapat selisih jumlah data, sehingga data diambil sesuai jumlah data terkecil, karena salah satu prasyarat model LSTM adalah jumlah data variabel harus sama. Data ini berasal dari sistem proses tahun 2018 hingga 2019. Jumlah data kemudian dibagi menjadi tiga bagian, yaitu untuk melatih model (data semester 1 2018 hingga semester kedua 2018), mengevaluasi model (data semester pertama 2019), dan terakhir digunakan untuk melakukan prediksi model (data semester kedua 2019). Jumlah data yang digunakan untuk melatih model adalah sebanyak 4592 buah data, untuk mengevaluasi adalah sebanyak 1968 buah data, dan untuk prediksi adalah sebanyak 3090 buah data.

3.3 Model LSTM

Algoritma *deep learning* ini disusun dengan bahasa pemrograman python dengan *library* NumPy, pandas, keras, dan scikit-learn, seaborn, scipy. Seluruh eksekusi kode python dilakukan pada server google collab.

Setelah algoritma *deep learning* disusun di program python, maka langkah selanjutnya adalah melatih model menggunakan *set data* tahun 2018 hingga semester pertama 2019 yang sudah dimodifikasi dengan metode interpolasi. Algoritma



Gambar 2. Hasil prediksi

Optimizer yang digunakan adalah SGD. Kinerja diukur dari nilai R^2 dengan nilai terbaik adalah nilai yang semakin mendekati 1 dan nilai RMSE yang semakin kecil. Hyperparameter yang digunakan dalam melatih model LSTM disajikan pada Tabel 3.

Tabel 3. *Hyperparameter* pelatihan

Hyperparameter	Nilai
Model Deep Learning	LSTM
Batch Sizes	Default (32)
Algoritma Optimizers	Stochastic Gradient Descent (SGD)
Learning Rate	Default (0.01)
Jumlah Epochs	200
Loss function	Mean Squared Error (MSE)
Hidden Unit	Default (128)
Activation Function	Default = Linear

3.3.1 Pelatihan dan evaluasi model

Model LSTM yang sudah dilatih memiliki nilai RMSE sebesar 1.4 dengan waktu *training* sebesar 321 detik. Model kemudian divalidasi untuk menghindari terjadi *overfitting* pada model. Hasil validasi model memiliki nilai RMSE sebesar 0,6389. Model ini dikatakan *fit* terhadap *training:test* data.

3.3.2 Prediksi model

Model yang telah dilatih dan divalidasi menggunakan set data tahun 2018 dan tengah semester tahun 2019 kemudian digunakan untuk memprediksi data pada tahun 2019 semester kedua. Hasil dari prediksi kemudian dievaluasi kembali menggunakan data kemurnian aktual proses. Hasil evaluasi menunjukkan nilai R^2 sebesar 0,74 dan nilai RMSE sebesar 0,56 dan disajikan secara grafis pada Gambar 2. Dengan nilai R^2 dan nilai RMSE tersebut dapat disimpulkan bahwa model tersebut *fit* terhadap prediksi sistem proses yang ada. Waktu eksekusi model LSTM hanya memakan waktu <2 detik dengan menggunakan server google colab. Model *deep learning* ini dapat dijadikan sebagai dasar dari penerapan *deep learning* pada industri kimia kolom distilasi.

4. KESIMPULAN

Variabel kondisi operasi proses memiliki hubungan tidak linier yang tidak bisa diwakilkan oleh korelasi Pearson. Variabel yang dipilih untuk melatih model adalah variabel "*press top column*" dan variabel "*upper temperature column*" berdasarkan *process knowledge*. Model LSTM dengan baik memprediksi set data pada semester 2 tahun 2019 dengan nilai RMSE sebesar 0,56 dan R^2 sebesar 0,74. Metode *deep learning* dapat diterapkan untuk pemodelan dinamik pada proses sistem distilasi

5. DAFTAR PUSTAKA

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250.

Choi, Y., Bhadriaju, B., Cho, H., Lim, J., Han, I. S., Moon, I., Kwon, J. S. Il, & Kim, J. (2023). Data-driven modeling of multimode chemical process: validation with a real-world distillation column. *Chemical Engineering Journal*, 457.

Esbensen, K. H., & Geladi, P. (2010). Principles of proper validation: use and abuse of re-sampling for validation. *Journal of Chemometrics*, 24(3–4), 168–187.

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: the role of machine learning. *IEEE Access*, 5, 20590–20616.

Karthikeyan, A., & Priyakumar, D. (2022). Artificial intelligence: machine learning for chemical sciences. *Journal of Chemical Sciences*.

Kwon, H., Oh, K. C., Choi, Y., Chung, Y. G., & Kim, J. (2021). Development and application of machine learning-based prediction model for distillation column. *International Journal of Intelligent Systems*, 36(5), 1970–1997.

Lee, J., Hong, S., Cho, H., Lyu, B., Kim, M., Kim, J., & Moon, I. (2021). Machine learning-based energy optimization for on-site SMR hydrogen production. *Energy Conversion and Management*, 244.

Mowbray, M., Vallerio, M., Perez-Galvan, C., Zhang, D., Del Rio Chanona, A., & Navarro-Brull, F. J. (2022). Industrial data science - A review of machine learning applications for chemical and process industries. In *Reaction Chemistry and Engineering* (Vol. 7, Issue 7, pp. 1471–1509). Royal Society of Chemistry.

Netrapalli, P. (2019). Stochastic gradient descent and its variants in machine learning. In *Journal of the Indian Institute of Science* (Vol. 99, Issue 2, pp. 201–213). Springer International Publishing.

Oda, Y., Kawamura, Y., & Fujishima, M. (2012). Energy consumption reduction by machining process improvement. *Procedia CIRP*, 4, 120–124.

Osuolale, F. N., & Zhang, J. (2016). Energy efficiency optimisation for distillation column using artificial neural network models. *Energy*, 106, 562–578.

Park, H., Kwon, H., Cho, H., & Kim, J. (2022). A framework for energy optimization of distillation process using machine learning-based predictive model. *Energy Science and Engineering*, 10(6), 1913–1924.

Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation functions in neural networks. In *International Journal of Engineering Applied Sciences and Technology* (Vol. 4).

Shewalkar, A. N. (2018). Comparison of RNN, LSTM and GRU on speech recognition data.