

# **PREDIKSI LAJU PERKEMBANGAN EKSPOR KACANG KEDELAI DI INDONESIA DENGAN METODE *LONG SHORT TERM MEMORY***

Syach Riyan M. A.<sup>1</sup>, Muhammad Ulwan F.<sup>2</sup>, dan Daniella Tasya E.<sup>3</sup>

Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia

E-mail: syach.riyan@ui.ac.id, muhammad.ulwan@ui.ac.id, daniella.tasya@ui.ac.id

## **Abstrak**

*Ekspor dan Impor merupakan faktor penting dalam pertumbuhan ekonomi di Indonesia. Tujuan dari penelitian ini adalah untuk melihat perkembangan laju salah satu bahan baku yang di ekspor oleh Indonesia yaitu kacang kedelai dengan penggunaan aplikasi machine learning. Data yang digunakan dalam penelitian ini adalah data sekunder ekspor kacang kedelai Indonesia yang diperoleh dari Data Dunia 2006-2016. Metode yang digunakan adalah Long Short Memory Term (LSTM) dengan melakukan pre-processing data, membangun model, dan evaluasi model. Hasil dari penelitian menunjukkan bahwa model dengan jumlah hidden layer LSTM = 3, batch size = 64, dan epoch = 75 dengan nilai RMSE sebesar 445.2332739403152 dan MAPE 26.8% adalah model terbaik dan yang paling sesuai dengan nilai yang sebenarnya.*

Kata kunci : Ekspor Kacang Kedelai, *Long Short Memory Term (LSTM)*, *Machine Learning*

## **Pendahuluan**

### **1. Latar Belakang**

Ekspor dan impor merupakan peranan penting dalam memenuhi kestabilan perekonomian di berbagai negara. Banyak negara yang sangat mengandalkan ekspor sebagai penghasilan utama kebutuhan untuk pendapatan nasional yang nantinya akan digunakan untuk keperluan negara termasuk Negara Indonesia. Banyak barang dan jasa yang diekspor dan diimpor oleh Indonesia, salah satunya adalah kacang kedelai.

Kacang kedelai merupakan bahan baku dasar yang sering dipakai oleh Negara Indonesia. Di Indonesia sendiri, buruh petani belum mampu untuk memenuhi kebutuhan kacang kedelai sehingga ada kalanya kacang kedelai diimpor dari luar negeri. Tetapi kadangkala, Indonesia mampu menghasilkan kacang kedelai sendiri tergantung dari keadaan pada saat itu. Namun menurut data dari Badan Pusat Statistik (BPS) mencatat bahwa impor kacang kedelai Indonesia pada semester 1 tahun 2020 mencapai 1,27 juta ton atau

sama dengan 510 juta *dollar AS* atau setara dengan Rp. 7,24 triliun (kurs Rp. 14,200,00).

Dapat dilihat dalam kurun waktu tahun 2020, Indonesia belum mampu menghasilkan kacang kedelai sendiri. Diharapkan kedepannya, Indonesia sudah mampu menghasilkan kacang kedelai untuk kebutuhan masyarakat sendiri dan mampu mengeksport kacang kedelai dengan jumlah banyak untuk memenuhi standar pendapatan nasional.

## 2. Rumusan Masalah

- Bagaimana perkembangan laju ekspor kacang kedelai di Indonesia?
- Bagaimana prediksi model *LSTM* untuk laju ekspor kacang kedelai di Indonesia?

## 3. Tujuan dan Manfaat

Tujuan paper ini dibuat untuk melihat perkembangan laju ekspor kacang kedelai di Indonesia dari tahun 2006-2016, serta membuat prediksi model laju ekspor kacang kedelai Indonesia untuk masa yang akan datang. Manfaat paper ini, diharapkan dapat membuat model terbaik untuk prediksi ekspor kacang kedelai di Indonesia.

## Data dan Batasan

### 1. Data

Data yang digunakan adalah data ekspor kacang kedelai Indonesia dari tahun 2006-2016. Data diambil dari *web resmi data.world.com* dengan kata kunci “Indonesia”. Data merupakan data

kuantitatif dengan jumlah 545 data yang tercatat.

### 2. Batasan

Batasan data yang digunakan adalah nilai ekspor kacang kedelai dari seluruh lahan yang ada di Indonesia. Nilai ekspor tersebut diakumulasi menjadi total nilai ekspor yang dihitung setiap beberapa kali dalam seminggu.

## Design dan Implementasi Analisis

### 1. Pre-processing

Proses *pre-processing* pada studi ini dilakukan penanganan *missing value*, pem dataset, serta normalisasi. *Missing value* dilakukan pertama dengan mengisi nilai pada kolom *AELast* (akumulasi nilai akhir ekspor) dengan nilai mediannya, kemudian karena ada beberapa data pada baris yang terisi pada waktu saja maka akan dihilangkan baris yang tidak memiliki nilai.

Selanjutnya pembagian dataset dilakukan dengan 38 data pada tahun 2016 sebagai data validasi dan 495 data pada tahun 2006-2015 sebagai data training. Setelah data dibagi selanjutnya data akan dinormalisasi pada studi ini menggunakan bantuan *MinMaxScaler* dari *sklearn*. Proses normalisasi ini bertujuan untuk mengubah interval data pada interval  $[0,1]$ . Adapun rumus normalisasi adalah sebagai berikut,

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x'$  = Nilai hasil normalisasi

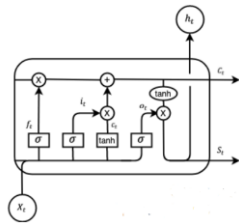
$x$  = Nilai yang akan dinormalisasi

$x_{min}$  = Nilai minimum dari data yang akan dinormalisasi

$x_{max}$  = Nilai maksimum dari data yang akan dinormalisasi

## 2. Membangun Model

Pada penelitian ini digunakan algoritma *Machine Learning Supervised* yaitu Metode *Long Short Term Memory (LSTM)*. Model *LSTM* dibangun dari sebuah *library* python yaitu *Keras*, sebuah *framework deep learning* di python. Model ini terdiri dari 3 atau 4 *hidden layers LSTM*. Model *LSTM* memiliki tiga *gate* yaitu,



Gambar 1. *Hidden Cell Memory LSTM*

### i. *Input Gate*

Pada bagian ini, *gate* berperan mengambil *output* sebelumnya dan *input* baru serta melewati mereka melalui lapisan *sigmoid*. *Gate* ini mengembalikan nilai 0 atau 1,

$$i_t = \sigma(W_i S_{t-1} + W_i X_t)$$

$W_t$  = Bobot pada *input gate*

$S_{t-1}$  = *State* sebelumnya

$X_t$  = *Input* pada waktu ke- $t$

$\sigma$  = Fungsi aktivasi (*sigmoid*)

Nilai gerbang *input gate* dikalikan dengan *output* dari lapisan kandidat ( $\hat{C}$ ) dengan rumus,

$$\hat{C} = \tanh(W_c S_{t-1} + W_c X_t)$$

$\hat{C}$  = *Intermediate cell state*

$W_c$  = Bobot pada *cell state*

### ii. *Forget Gate*

Pada bagian ini, *gate* berperan mengontrol sejauh mana nilai tetap di dalam sel memori. Dengan rumus,

$$f_t = \sigma(W_f S_{t-1} + W_f X_t)$$

$W_f$  = Bobot pada-*i forget gate*

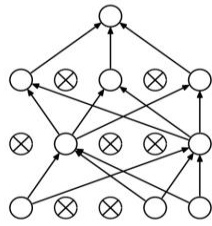
$$C_t = i_t * \hat{C}_t + f_t * C_{t-1}$$

### iii. *Output Gate*

$$o_t = \sigma(W_o S_{t-1} + W_o X_t)$$

$$h_t = o_t \tanh(c_t)$$

Pada setiap *layer* juga digunakan fungsi aktivasi *ReLU* yang merupakan fungsi aktivasi yang paling populer dan banyak diterapkan dalam pembangunan model *neural network* (Ramachandran, Zoph, & Le, 2017). Semua *input* nilai negatif pada fungsi aktivasi *ReLU* akan menjadi nilai 0. Selanjutnya masing-masing *layers LSTM* memiliki *Dropout*. *Dropout* merupakan cara yang efisien untuk menghindari *overfitting*. *Dropout* dilakukan dengan cara mematikan (*dropping out*) dari lapisan tersembunyi (*hidden layer*) maupun lapisan yang tampak (*visible layer*).



Gambar 2. *Neural Network* setelah diaplikasikan *Dropout*

Setelah model dibuat selanjutnya model di compile menggunakan *ADAM*. *ADAM* adalah algoritma yang digunakan untuk mengoptimasi gradien pada *neural network* berdasarkan data *train*. Metode ini lebih mudah diimplementasikan, efisien secara komputasi, memerlukan kebutuhan memori yang kecil, dan sesuai untuk masalah yang memiliki banyak data maupun parameter (Kingma & Ba, 2015).

### 3. Evaluasi Model

Pada penelitian ini evaluasi model dilakukan dengan melihat nilai dari *RMSE* (*Root Mean Square Error*) dan *MAPE* (*Mean Absolute Percentage Error*). *MAPE* merupakan rata-rata diferensiasi absolut antara nilai prediksi dan nilai sebenarnya, yang dinyatakan sebagai persentase nilai sebenarnya. *MAPE* digunakan untuk menghitung persentase kesalahan antara nilai prediksi dan nilai sebenarnya. Ada berbagai ukuran akurasi dalam prediksi, diantaranya adalah *MAPE* dan *RMSE*. Berikut adalah rumusnya,

#### a) *Root Mean Square Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n} \sum_i^n (\tilde{y} - y_i)^2}$$

$\tilde{y}$  = Nilai hasil prediksi

$y_i$  = Nilai sebenarnya

$n$  = Jumlah data

#### b) *Mean Absolute Percentage Error (MAPE)*

$$MAPE = 100 \times \frac{1}{n} \sum_i^n \left| \frac{\tilde{y} - y_i}{y_i} \right|$$

$\tilde{y}$  = Nilai hasil prediksi

$y_i$  = Nilai sebenarnya

$n$  = Jumlah data

## Hasil dan Analisis

Pada studi ini sebelum percobaan dilakukan, ada beberapa parameter yang berpengaruh dalam model seperti jumlah *hidden layer* (jumlah lapisan *LSTM* pada model), nilai *epoch* (*hyperparameter* yang menentukan berapa kali algoritma pembelajaran akan bekerja mengolah seluruh dataset *training*), dan juga *batch size* (jumlah sampel data yang disebarkan ke *neural network*) dengan nilai yang berbeda akan menghasilkan akurasi yang berbeda juga. Oleh karenanya ada beberapa kasus yang dibuat seperti dibawah ini,

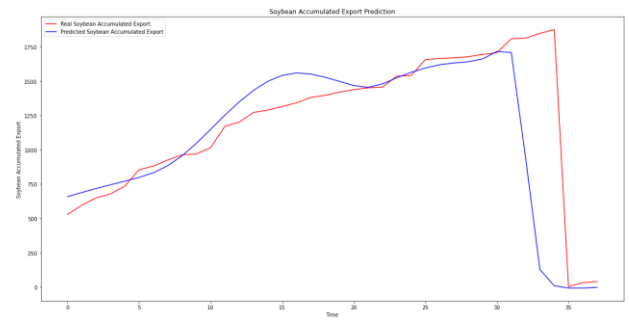
	1	2
P1 ( <i>Hidden Layer</i> )	3	4
P2 ( <i>Batch size</i> )	32	64
P3 ( <i>Epoch</i> )	50	75

Tabel. 1. Rancangan Kasus Percobaan

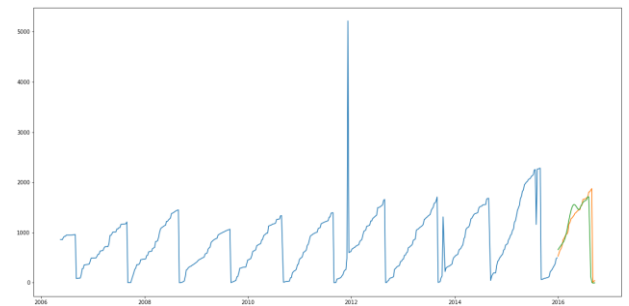
Dari kasus yang dibuat, dihasilkan nilai *RMSE* dan *MAPE* sebagai berikut,

	RMSE	MAPE
P1,1, P2,1,P3,1	496.8631830921288	36.7%
P1,1, P2,1,P3,2	428.5314351543482	30.4%
P1,1, P2,2,P3,1	389.35711509779685	111.3%
<b>P1,1, P2,2,P3,2</b>	<b>445.2332739403152</b>	<b>26.8%</b>
P1,2, P2,1,P3,1	439.82097317935126	46.1%
P1,2, P2,1,P3,2	415.6541212730986	40.8%
P1,2, P2,2,P3,1	311.7555993716864	100.2%
P1,2, P2,2,P3,2	490.2672814844983	37.4%

Tabel 2. Hasil Nilai *Error* Setiap Kasus  
Terlihat pada tabel 2 bahwa nilai *MAPE* terkecil adalah pada kasus **P<sub>1,1</sub>, P<sub>2,2</sub>, P<sub>3,2</sub>** yakni saat jumlah *hidden layer* = 3, *batch size* = 64, dan *epoch* = 75 dengan nilai *RMSE* sebesar 445.2332739403152 dan *MAPE* 26.8%. Berikut grafik dari model tersebut,



Gambar 3. Grafik Perbandingan Nilai Prediksi dan Valid



Gambar 4. Grafik Perbandingan Nilai Prediksi dan Valid dengan Nilai Sebelumnya

## Kesimpulan

Hasil dari visualisasi data dapat dilihat bahwa perkembangan laju ekspor kacang kedelai di Indonesia mengalami peningkatan yang tidak terlalu signifikan pada setiap tahunnya dan uniknya pada laju ekspor terlihat adanya titik waktu dimana nilai ekspor mengalami penurunan yang lumayan ekstrim. Hal ini dikarenakan kebutuhan masyarakat Indonesia terhadap kacang kedelai masih tinggi. Kemudian untuk beberapa kasus model *LSTM* yang sudah didapatkan nilai prediksinya, didapatkan model dengan jumlah *hidden layer LSTM* = 3, *batch size* = 64, dan *epoch* = 75 dengan nilai *RSME* sebesar 445.2332739403152 dan *MAPE* 26.8% adalah percobaan model *LSTM* terbaik dan

visualisasi datanya mendekati dengan nilai sebenarnya. Meskipun nilai *error* yang dihasilkan masih dirasa belum minimum namun pada visualisasi data prediksi tetap bisa menggambarkan nilai yang sebenarnya dimana pada tahun pertama dan seterusnya mengalami peningkatan namun pada waktu di pertengahan tahun ada waktu yang menyebabkan nilai ekspor menurun drastis.

## Saran

- Pemerintah perlu membuka lahan baru dan menambah lapangan kerja untuk dapat mengembangkan ekspor kacang kedelai di Indonesia.
- Pemanfaatan data laju ekspor kacang kedelai dinilai sebagai pemicu dalam

mengembangkan ekspor di Indonesia untuk memajukan perekonomian Indonesia dan juga untuk memenuhi kebutuhan masyarakat Indonesia.

- Pada data tahun sebelumnya dan data prediksi, ada waktu dimana nilai ekspor kacang kedelai akan mengalami penurunan drastis, sehingga dapat dilakukan pengamatan lebih lanjut mengenai kejadian ini sehingga laju ekspor kacang kedelai semakin bisa dikendalikan dan memberikan keuntungan yang maksimal.
- Untuk menghasilkan prediksi data yang lebih akurat dapat dikumpulkan data yang lebih lengkap dan mencoba parameter atau pun model lainnya.

## Daftar Pustaka

- [1] Bea Cukai Meulaboh. *Pengenalan Ekspor*. Diakses tanggal 24 Okt. 2021, dari <http://bcmeulaboh.beacukai.go.id/>
- [2] Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep learning*. Nature.
- [3] Boiko, Sergii. 2018. *Practical Machine Learning for Solving Real World Problems*. Railsware. Diakses pada 23 Maret 2021, dari <https://railsware.com/blog/practical-machine-learning-for-solving-real-world-problems/>
- [4] L. Skovajsová, “Long Short-Term Memory Description and its Application in Text Processing.”
- [5] R. Vinayakumar, K. P. Soman, and P. Poornachandran, “Long short-term memory based operation log anomaly detection,” 2017 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2017, vol. 2017-Janua, pp. 236–242, 2017.
- [6] G. Mathisen, “Forecasting Multivariate Time Series Data Using Neural Networks Sigurd Øyen,” no. June, 2018.
- [7] X. Qu, X. Kang, Z. Chao, J. Shuai, and X. Ma, “Short-term prediction of wind power based on deep Long Short-Term Memory,” Asia-Pacific Power Energy Eng. Conf. APPEEC, vol. 2016-Decem, pp. 1148–1152, 2016.
- [8] Diakses pada tanggal 24 Oktober 2021 <https://money.kompas.com/read/2021/01/03/134256526/mengapa-indonesia-begitu-bergantung-pada-kedelai-impor-dari-as?page=all>
- [9] H. Kusdarwati & S. Handoyo, “System for prediction of non stationary time series based on the wavelet radial bases function neural network model,” Int. J. Electr. Comput. Eng., vol. 8, no. 4, pp. 2327–2337, 2018.
- [10] Wiranda, Laras & Mujiono, Sadikin. (2019) “PENERAPAN LONG SHORT TERM MEMORY PADA DATA TIME SERIES UNTUK MEMPREDIKSI PENJUALAN PRODUK PT. METISKA FARMA”. Universitas Mercu Buana.
- [11] Rizal, Aminuddin. (2020). *Tahapan Desain dan Implementasi Model Machine Learning untuk Sistem Tertanam*. Universitas Komputer Indonesia

## Lampiran

- *Data*

	Date	UBCurr	UBLast	AECurr	AELast	TCCurr	TCLast	% change	next year
0	2006-05-18	113.6	43.8	944.7	858.8	1058.0	903.0	17.250166	NaN
1	2006-05-25	58.9	43.8	1010.7	858.8	1070.0	903.0	18.502105	NaN
2	2006-06-01	57.8	43.8	1011.8	858.8	1070.0	903.0	18.502105	NaN
3	2006-06-08	57.0	4.8	1012.6	900.8	1070.0	906.0	18.109541	0.0
4	2006-06-15	61.5	4.8	1012.6	924.1	1074.0	929.0	15.631392	0.0

Cuplikan Data

- *Modelling*

Untuk proses lengkap membangun model, dapat diakses di link github:  
<https://github.com/syachmhrd/DAS/tree/main/Percobaan>

```
INPUT DATA dan Packages

In [1]: import numpy as np, pandas as pd, matplotlib.pyplot as plt

In [2]: from matplotlib.pyplot import rcParams
rcParams['figure.figsize']=20,10
from keras.models import Sequential
from keras.layers import LSTM,Dropout,Dense
from sklearn.preprocessing import MinMaxScaler

In [3]: data = pd.read_excel('https://github.com/syachmhrd/DAS/blob/main/Soybean%20Indo%20FIX.xlsx?raw=true')

PreProcessing

In [4]: data.head()
Out[4]:
```

	Date	UBCurr	UBLast	AECurr	AELast	TCCurr	TCLast	% change	next year
0	2006-05-18	113.6	43.8	944.7	858.8	1058.0	903.0	17.250166	NaN
1	2006-05-25	58.9	43.8	1010.7	858.8	1070.0	903.0	18.502105	NaN
2	2006-06-01	57.8	43.8	1011.8	858.8	1070.0	903.0	18.502105	NaN
3	2006-06-08	57.0	4.8	1012.6	900.8	1070.0	906.0	18.109541	0.0
4	2006-06-15	61.5	4.8	1012.6	924.1	1074.0	929.0	15.631392	0.0

```

In [5]: data.info()

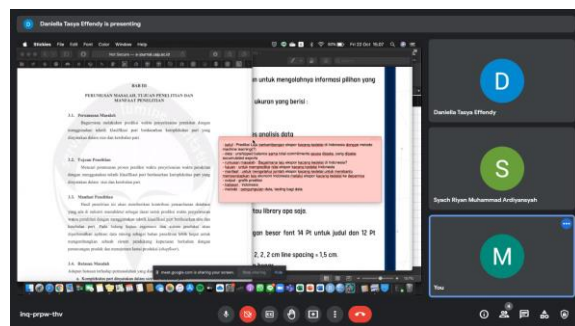
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 587 entries, 0 to 586
Data columns (total 9 columns):
```

Cuplikan Github

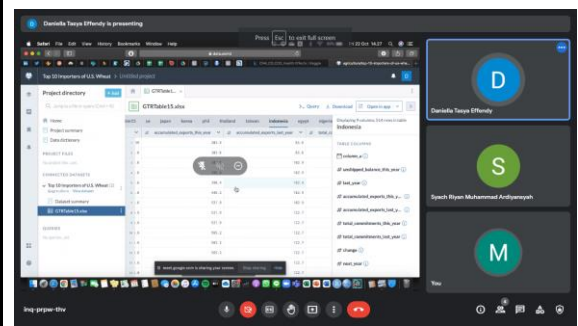
- *Dokumentasi*



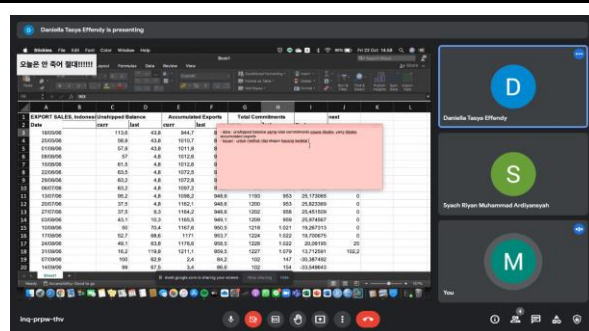
Dokumentasi 1. Anggota Kelompok



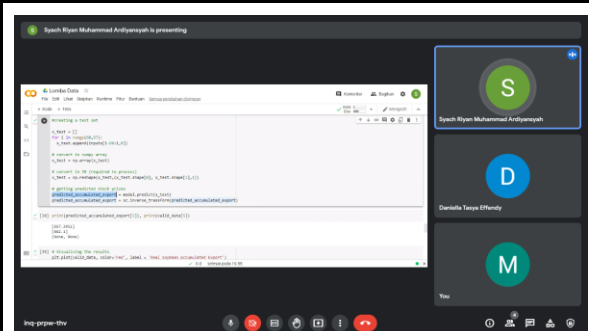
Dokumentasi 2. Menentukan Ide, Mencari Sumber



Dokumentasi 3. Mencari Data



Dokumentasi 4. Mencari data



Dokumentasi 5. Membangun Model



