

APLIKASI KLASIFIKASI KUALITAS DAN PREDIKSI USI SIMPAN BERAS BERBASIS *DATASET ELECTRONIC NOSE* MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*

Indah Mayang Sari¹, Dedy Rahman Wijaya², Wahyu Hidayat³

^{1,2,3}Universitas Telkom, Bandung

inmaysasi@student.telkomuniversity.ac.id¹, dedyrw@telkomuniversity.ac.id²,

wahyuhidayat@telkomuniversity.ac.id³

Abstrak

Beras merupakan makanan pokok masyarakat di Asia, khususnya di Indonesia. Cadangan beras yang optimal menyediakan kebutuhan pangan yang cukup. Masalah mendasar beberapa tahun terakhir adalah bahwa kualitas beras telah memburuk dalam beberapa tahun terakhir, sehingga mengakibatkan kerugian. Dalam hal ini aplikasi ini dibangun untuk melakukan klasifikasi kualitas dan prediksi usia simpan beras menggunakan dataset berbasis *electronic nose* dengan memanfaatkan kemampuan algoritma *Naïve Bayes Classifier* dan *Bayesian Ridge*. Klasifikasi menggunakan 4 model *naïve bayes* yaitu *Gaussian Naïve Bayes* dengan akurasi 82%, *Multinomial Naïve Bayes* memperoleh akurasi 97%, *Complement Naïve Bayes* memperoleh akurasi 98%, dan *Bernoulli Naïve Bayes* memperoleh akurasi 54%. Sedangkan *Bayesian Ridge* memperoleh *Root Mean Square Error* (RMSE) 3,58 dan *R-Squared* sebesar 0,72. Model klasifikasi ini akan ditampilkan dalam bentuk web yang dapat memprediksi data baru dengan atribut yang telah ditentukan. Diharapkan penelitian ini dapat membantu beberapa pihak terkait yang berpengaruh untuk produksi beras dalam mengambil keputusan.

Kata Kunci : Beras, *Naïve Bayes*, *Gaussian*, *Multinomial*, *Complement*, *Bernoulli*, *Bayesian Ridge*, Klasifikasi, Regresi

Abstract

Rice is a staple cuisine for people in Asia, particularly in Indonesia. Rice reserves that are optimal provide sufficient dietary needs. The fundamental issue is that rice quality has deteriorated in recent years, resulting in losses. The traditional method for determining rice quality is to utilize human senses to detect the aroma of rice and examine the texture of the rice. In this case, this application is built to classify quality and predict shelf life using an electronic nose-based dataset by utilizing the capabilities of the Naïve Bayes Classifier and Bayesian Ridge algorithms. The classification uses 4 naïve Bayes models, namely Gaussian Naïve Bayes with 82% accuracy, Multinomial Naïve Bayes with 97% accuracy, Complement Naïve Bayes with 98% accuracy, and Bernoulli Naïve Bayes with 54% accuracy. Meanwhile, Bayesian Ridge obtained a Root Mean Square Error (RMSE) of 3.58 and an R-Squared of 0.72. This classification model will be displayed in a web form that can predict new data with predetermined attributes. It is hoped that this research can help several stakeholder that have an influence on rice production in decision making

Keywords : *Rice*, *Naïve Bayes*, *Gaussian*, *Multinomial*, *Complement*, *Bernoulli*, *Bayesian Ridge*, Classification, Regression

1. Pendahuluan

Bagi Indonesia, beras merupakan komoditas pangan yang sangat strategis karena makanan pokok utama bagi sebagian besar penduduknya. Produksi pada 2019, beras konsumsi pangan penduduk sebesar 31,31 juta ton beras[1]. Ketersediaan pangan yang cukup harus didukung oleh adanya surplus beras sebagai cadangan pangan. Pengelolaan cadangan beras ini diamanatkan pemerintah kepada Perusahaan Umum Badan Urusan Logistik atau BULOG [2].

Namun masih banyak ditemukan beras turun mutu dan tidak layak konsumsi beberapa tahun terakhir. Contoh kasus ditemukan 6.800 Ton yang berlokasi di Bulog Drive Sumatra Selatan dan Bangka Belitung pada 13 Februari 2019 [3], sehingga dilakukan sortasi dan pemisahan unit yang berbeda untuk menghindari terkontaminasinya beras baik [3]Hal ini disebabkan karena melebihi batas penyimpanan di gudang, yaitu batas maksimal 4 bulan. Bukti lapangan beras-beras yang ditemukan telah melebihi batas penyimpanan yaitu sekitar 5 bulan hingga 2 tahun lama penyimpanan di gudang [4].

Penyimpanan merupakan tahap yang menentukan dalam menjamin ketersediaan beras berkualitas[5]. Selama penyimpanan, beras mengalami penyusutan kualitas dan kuantitas yang disebabkan oleh perubahan fisik, kimia, dan biologi [5]. Maka dari itu pola produksi yang baik yang didukung dengan pendistribusian yang optimal dapat menghindari terjadinya kadaluarsa.

Dengan demikian akan dikembangkan prototype aplikasi berbasis dataset electronic nose dengan memanfaatkan algoritma Naive Bayes untuk klasifikasi kualitas beras dan Bayesian Ridge untuk prediksi usia simpan beras. Aplikasi ini diharapkan dapat membantu pihak yang bersangkutan dalam pengambilan keputusan.

2. Metode Penelitian

Metode pengembangan perangkat lunak yang digunakan pada pengerjaan proyek akhir ini adalah dengan menggunakan SDLC *Waterfall* . Berikut gambaran tahapan – tahapan SDLC *Waterfall*.



Gambar 1 Metode Penelitian

Dari gambar 1 dapat dijabarkan mengenai metode pengerjaan yang digunakan saat pengerjaan proyek akhir. SDLC waterfall digunakan karena pada pengembangan aplikasi ini menggunakan waktu yang relatif pendek [5].

1. Analisis

Pada tahap ini bertujuan untuk memahami perangkat lunak yang akan dibangun dengan mencari data-data yang diperlukan oleh penulis dengan cara wawancara, serta tinjauan pustaka atau studi literatur dengan mencari referensi buku, web, jurnal yang berhubungan dengan perangkat lunak yang akan dibangun

2. Desain

Pada tahapan ini penulis menggunakan bahasa pemrograman python untuk membangun perangkat lunak atau aplikasi, menggunakan tools BPMN untuk menggambarkan proses bisnis yang sedang berjalan (As-Is) atau yang akan berjalan (To-Be) dan merancang database menggunakan ERD untuk menentukan atribut serta entitas-nya.

3. Implementasi

Proses implementasi pada proyek akhir ini akan menggunakan bahasa pemrograman python karena bahasa pemrograman python dapat lebih mudah menggunakan library yang memudahkan proses prediksi data menggunakan Machine Learning. Serta antarmuka yang akan digunakan adalah dengan menggunakan framework flask yang berbasis web.

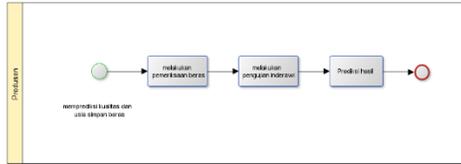
4. Pengujian

Aplikasi proyek akhir ini akan diuji dengan melakukan pengujian metode black-box. Akan tetapi tidak menutup kemungkinan aplikasi ini akan diuji dengan metode white-box. Karena dalam proses Machine Learning terdapat hal – hal yang harus diuji untuk memaksimalkan performa aplikasi.

3. Hasil dan Pembahasan

3.1 Gambaran Sistem Saat Ini

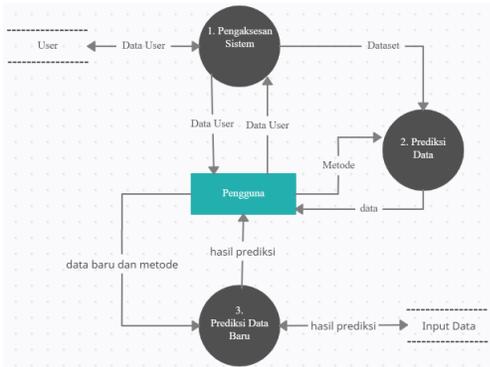
Berikut menunjukkan sistem yang sedang berjalan dengan menggunakan Direct Method. Direct Method adalah metode langsung yang paling umum digunakan. Dalam hal ini, bagian produsen menggunakan metode konvensional atau direct method yang memanfaatkan indera manusia untuk mengetahui kualitas beras.



Gambar 2 Sistem Saat Ini

3.2 Gambaran Metode Usulan

Pada proses ini data beras yang akan diolah masuk ke dalam proses input data beras untuk dilakukan pembersihan data, setelah itu data yang telah dibersihkan akan masuk ke dalam tahap normalisasi data. Jika tahap normalisasi berhasil, maka data akan masuk ke dalam tahap prediksi dan klasifikasi data. Pada proses ini data akan menghasilkan hasil akhir yaitu klasifikasi kualitas dan prediksi usia simpan beras yang dapat di lihat oleh pengguna pada tampilan web.

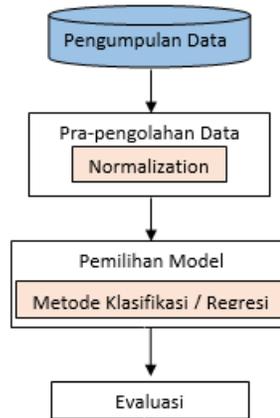


Gambar 3 Metode Usulan

3.3 Ilustrasi Perangkat Lunak

Aplikasi ini memanfaatkan kemampuan *Machine Learning* menggunakan algoritma *Naive Bayes* untuk klasifikasi kualitas beras dan *Bayesian Ridge* untuk memprediksi usia simpan beras. Dengan menggunakan data *electronic nose* sebagai dataset untuk data training pada *machine learning*, maka sistem akan mengeluarkan hasil klasifikasi kualitas dan prediksi usia simpan. Dalam proses pengolahan data hingga terbentuk model klasifikasi dan regresi

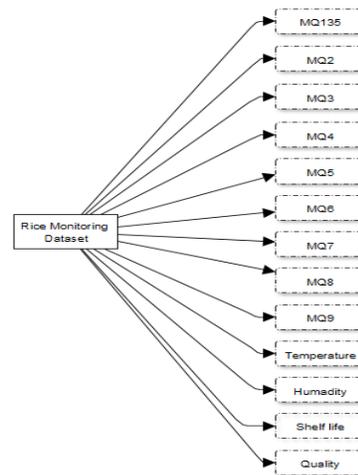
terdapat proses seperti pada gambar berikut :



Gambar 4 Ilustrasi Perangkat Lunak

a. Dataset

Proses pengumpulan data merupakan pengelompokan data yang akan digunakan untuk proses machine learning. Data yang diambil merupakan dataset electronic nose yang memiliki berbagai fitur dan juga label. Adapun fitur - fitur yang akan digunakan yaitu fitur MQ135, MQ2, MQ3, MQ4, MQ5, MQ6, MQ7, MQ8, MQ9, Humidity, Temperature, sedangkan label shelf life untuk prediksi dan label Quality untuk klasifikasi. Dataset yang digunakan sebanyak 48,846 data dengan 22,804 data kadaluarsa dan 26,042 data tidak kadaluarsa. Gambar 15 menunjukkan struktur fitur dan label pada dataset yang akan digunakan :



Gambar 5 Struktur Data

b. Pra – Pengolahan Data

Pada tahap pra-pengolahan data digunakan normalisasi untuk mengubah data agar memiliki

distribusi yang serupa. Pada kasus ini penulis menerapkan metode min – max scaler yang bekerja dengan scaling data atau menyesuaikan data dalam rentang tertentu (rentang nilai minimum hingga nilai maksimum), dengan rentang yang biasa digunakan adalah 0 hingga 1. Sebagai contoh data nilai 15 dengan nilai fitur 15, 20, 25, 30 yang akan dinormalisasi dengan contoh sebagai berikut :

$$x' = \frac{15 - 15}{30 - 15}$$

$$x' = \frac{0}{15}$$

$$x' = 0$$

Dari hasil tersebut dapat disimpulkan nilai sebelum dinormalisasi 15 dan telah dinormalisasi menjadi 0.

c. Pemodelan

Pada tahap learning atau pembelajaran, dataset akan dilatih dan diuji menggunakan metode Naïve Bayes Classifier untuk kasus klasifikasi dan Bayesian Ridge untuk kasus regresi. Pada pemodelan klasifikasi, Naïve Bayes memiliki sifat yang berakar pada teorema bayes, yaitu metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode ini memiliki beberapa model yaitu Gaussian Naïve Bayes, Multinomial Naïve Bayes, Complement Naïve Bayes, dan Bernoulli Naïve Bayes. Setiap model memiliki asumsi tentang distribusi probabilitas target terhadap prediktor yang berbeda. Dataset yang digunakan akan dibagi menjadi data latih (training) dan data uji (testing). Dalam pemodelan, performansi akan ditingkatkan menggunakan grid search dengan parameter terbaik dari setiap algoritma.

d. Evaluasi

Pada tahap evaluasi, setiap model akan diukur performannya. Untuk kasus klasifikasi, pemodelan akan diukur menggunakan *Confusion Matrix*, *Receiver Operating Characteristic Curve*. Untuk kasus regresi akan diukur menggunakan *Mean Square Error* dan *Root Mean Square Error*.

3.4 Implementasi

3.4.1 Implementasi Machine Learning

a. Pra- Pengolahan Data

Pada tahap *pre-processing*, data akan melakukan proses pemeriksaan data yang null, untuk kemudian dinormalisasi agar memiliki nilai distribusi yang sama. Data akan diambil dari dataset dan diperiksa apakah data mengandung data null. Pada pemeriksaan data null, hasil menunjukkan bahwa data null tidak ada, yang berarti dataset terisi semua. Pada pemodelan regresi, Usia (minggu) akan menjadi target yang akan diprediksi.

```
# 1. Get Data Training
dftrain = pd.read_excel('dataatrain.xlsx')
X_train = dftrain[['MQ135', 'MQ2', 'MQ3', 'MQ4', 'MQ5', 'MQ6', 'MQ7', 'MQ8', 'MQ9', 'Temperature', 'Humidity']]
y_train = dftrain['Usia (minggu)']

# 2. Get Data Testing
dftest = pd.read_excel('datatest.xlsx')

X_test = dftest[['MQ135', 'MQ2', 'MQ3', 'MQ4', 'MQ5', 'MQ6', 'MQ7', 'MQ8', 'MQ9', 'Temperature', 'Humidity', ]]
y_test = dftest['Usia (minggu)']

dftrain.isnull().sum()
output :
```

MQ135	0
MQ2	0
MQ3	0
MQ4	0
MQ5	0
MQ6	0
MQ7	0
MQ8	0
MQ9	0
Temperature	0
Humidity	0
Usia (minggu)	0

```
dtype: int64
```

b. Pemodelan

Pada tahap pemodelan akan dilakukan pengujian menggunakan 4 model naïve bayes yaitu Gaussain, Multinomial, Complement dan Bernoulli. Dalam meningkatkan performa model, akan diterapkan metode grid search untuk mengetahui parameter terbaik dari tiap model. Implementasi pemodelan ditunjukkan sebagai berikut :

```
np.random.seed(999)
cv_method = RepeatedStratifiedKFold(n_splits=10, n_repeats=3, random_state=999)

# Gaussian NB
classifier1 = GaussianNB()
params_nb1 = {'var_smoothing': np.logspace(0, -9, num=100)}
gs_NB = GridSearchCV(estimator=classifier1, param_grid=params_nb1, verbose=1, cv=5, scoring='accuracy')
gs_NB.fit(x1, y_train)
y_pred_gsn = gs_NB.predict(x2)

# Bernoulli NB
classifier2 = BernoulliNB()
params_nb2 = {'alpha': [0.01, 0.1, 0.5, 1.0, 10.0],}
brn_NB = GridSearchCV(estimator=classifier2, param_grid=params_nb2, verbose=1, cv=5, scoring='accuracy')
brn_NB.fit(x1, y_train)
y_pred_brn = brn_NB.predict(x2)

# Multinomial NB
```

```
classifier3 = MultinomialNB()
params_nb3 = {'alpha': [0.01, 0.1, 0.5, 1.0, 10.0]}
mlt_NB = GridSearchCV(estimator=classifier3, param_grid=params_nb3, verbose=1, cv=5, scoring='accuracy')
mlt_NB.fit(x1, y_train)
y_pred_mlt = mlt_NB.predict(x2)

# Complement NB
classifier4 = ComplementNB()
params_nb4 = {'alpha': [0.01, 0.1, 0.5, 1.0, 10.0]}
cmp_NB = GridSearchCV(estimator=classifier4, param_grid=params_nb4, verbose=1, cv=5, scoring='accuracy')
cmp_NB.fit(x1, y_train)
y_pred_cmp = cmp_NB.predict(x2)
```

c. Evaluasi

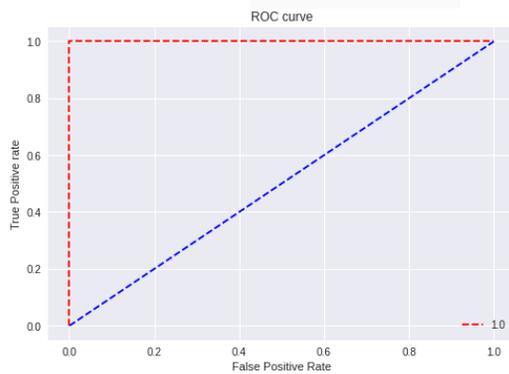
Pada evaluasi model klasifikasi, akan ditampilkan menggunakan confusion matrix, ROC Value, dan Classification Report Berikut implementasi dan data evaluasi. Table 12 menunjukkan hasil confusion matrix dari hasil klasifikasi menggunakan keempat model sebagai berikut :

		Expired	Non-expired
Gaussian NB	Precision	0,73	1,00
	Recall	1,00	0,66
	F-Measure	0,85	0,80
	Accuracy	0,82	
Bernoulli NB	Precision	1,00	0,52
	Recall	0,01	1,00
	F-Measure	0,02	0,69
	Accuracy	0,52	
Multinomial NB	Precision	1,00	0,95
	Recall	0,94	1,00
	F-Measure	0,97	0,97
	Accuracy	0,97	

Complement NB	Precision	0,96	1,00
	Recall	1,00	0,96
	F-Measure	0,98	0,98
	Accuracy	0,98	

Berikut tahap implementasi ROC Curve untuk melihat Area Under Curve dari setiap metode :

```
# GaussianNB
plt.plot(fpr1, tpr1, lines
tyle='--
',color='red', label=auc_s
core1)
plt.plot(p_fpr, p_tpr, lin
estyle='--', color='blue')
# title
plt.title('ROC curve')
# x label
plt.xlabel('False Positive
Rate')
# y label
plt.ylabel('True Positive
rate')
plt.legend(loc='best')
plt.savefig('ROC',dpi=300)
plt.show()
```

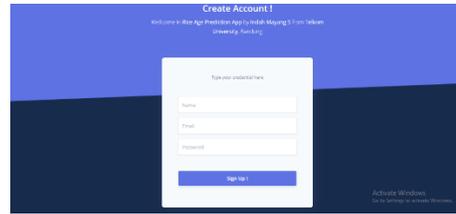


Gambar 6 ROC Curve

3.4.2 Implementasi Desain Antarmuka

a. Halman Registrasi

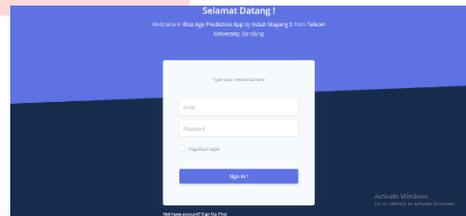
Pada halaman Sign in user akan berinteraksi dengan tampilan halaman log on, yang dimana user diwajibkan untuk melakukan pendaftaran akun.



Gambar 7 Halaman Registrasi

b. Halaman Log In

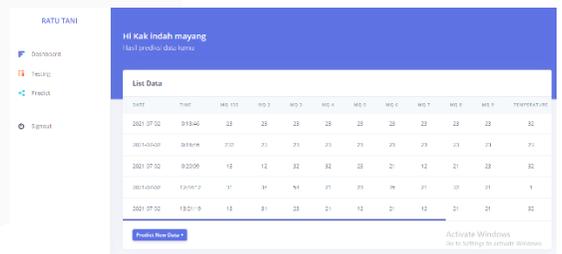
Pada halaman ini, user akan berinteraksi dengan tampilan log in, yang dimana mengharuskan user untuk memasukan alamat email dan password yang sudah terdaftar.



Gambar 8 Halaman Log In

c. Halaman Dashboard

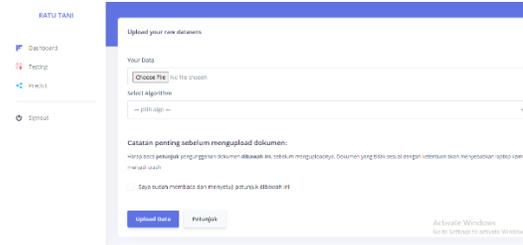
Pada halaman dashboard, user dapat melihat informasi akses penggunaan website dan laporan akhir penggunaan data pada website.



Gambar 9 Halaman Dashboard

d. Halaman Upload Data

Pada halaman upload data, user dapat melihat form pengisian input data yang ingin digunakan.



Gambar 10 Halaman Upload Data

e. Halaman Hasil Prediksi

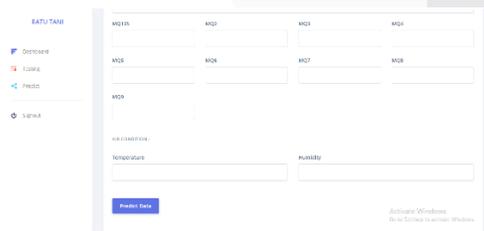
Pada halaman hasil prediksi user dapat melihat hasil prediksi dari dataset yang digunakan.



Gambar 11 Halaman Hasil Prediksi

f. Halaman Prediksi Data Baru

Pada halaman prediksi data baru, user dapat melakukan input data baru yang akan diprediksi tanpa melakukan upload dataset. Tampilan berupa form input atribut data.



Gambar 12 Halaman Prediksi Data Baru

4. Kesimpulan

Dalam proses pengembangan aplikasi klasifikasi kualitas dan prediksi usia simpan beras ini dapat disimpulkan sebagai berikut :

- 1) Dalam proyek akhir ini, dibangun algoritma *machine learning Naïve Bayes Classifier*

untuk memprediksi kualitas beras. *Naïve Bayes* memiliki beberapa macam metode yaitu, *Gaussian Naïve Bayes*, *Bernoulli Naïve Bayes*, *Multinomial Naïve Bayes* dan *Complement Naïve Bayes*. Klasifikasi beras memiliki akurasi yang berbeda, *Gaussian* memperoleh 82%, *Bernoulli* memperoleh 52%, *Multinomial* 97%, dan *Complement* 98%. Klaisifikasi ini menggunakan data *electronic nose* yang telah dinormalisasi menggunakan *min – max scaler*.

- 2) Dalam proyek akhir ini, dibangun algoritma *machine learning Bayesian Ridge* menggunakan dataset *electronic nose* untuk memprediksi usia simpan beras. Prediksi usia simpan memperoleh *R – Squared* sebesar 0,72 dan *Root Mean Square Error* 3,58.
- 3) Dalam Proyek akhir ini sudah dikembangkan aplikasi berbasis web yang menjalankan algoritma *machine learning Naïve Bayes Classifier* dan *Bayesian Ridge*. Berdasarkan hasil pengujian, aplikasi ini bisa digunakan oleh pihak – pihak terkait.

Referensi

- [1] Badan Pusat Statistik, “Luas Panen dan Produksi Padi di Indonesia 2019,” *Ber. Resmi Stat.*, no. 16, p. 2, 2020, [Online]. Available: <https://www.bps.go.id/pressrelease/download.html?nrbvfeve=MTc1Mg%3D%3D&sdfs=ldjfdifsdjfkfahi&twoadfnoarfeauf=MjAyMCOwMi0yNSAxMzoyNjozMg%3D%3D>.
- [2] Perusahaan Umum Badan Urusan Logistik, “Ketahanan Pangan BULOG,” *BULOG*, 2014. <http://www.bulog.co.id/ketahananpangan.php>.
- [3] Redaksi, “Dicari, Solusi Menguras Beras Bulog,” 2019. [Online]. Available: <http://agroindonesia.co.id/2019/03/dicari-solusi-menguras-beras-bulog/>.
- [4] Kompas.com, “20.000 Ton Beras Bulog yang Rusak akan Dijual Murah,” *Kompas*, 2019. <https://nasional.kompas.com/read/2019/12/04/20383761/20000-ton-beras-bulog-yang-rusak-akan-dijual-murah>.
- [5] Ratnawati, M. Djaeni, and D. Hartono, “Perubahan kualitas beras selama penyimpanan,” *Pangan*, vol. 22, no. 3, pp. 199–208, 2013.

