



Prediksi Saldo Produksi Hasil Ternak Kabupaten Blitar Menggunakan Regresi Linier Berganda

Ramadhan Paninggalih^{1, a)}, Bowo Nugroho^{2, b)}, dan
Mahardika Inra Takaendengan^{3, c)}

^{1,2}Program Studi Informatika, Institut Teknologi Kalimantan, Jalan Soekarno Hatta KM.15, Balikpapan

³Program Studi Sistem Informasi, Universitas Sam Ratulangi, Bahu, Malalayang, Manado

^{a)}Penulis korespondensi: ramadhanpaninggalih@lecturer.itk.ac.id

^{b)}bowo.nugroho@lecturer.itk.ac.id

^{c)}mtakaendengan@dosenlb.unsrat.ac.id

Abstrak. Penyesuaian masyarakat kota Blitar terhadap perkembangan teknologi 4.0 di Indonesia dengan kecepatan dan ketepatan dalam penguatan persaingan bisnis, khususnya di bidang sentra peternakan Indonesia merupakan sebuah kewajiban. Ketepatan dan kecepatan informasi mengenai saldo produksi hasil ternak mempengaruhi optimalisasi kualitas dan kuantitas sektor Usaha mikro kecil menengah (UMKM) peternakan di Kota Blitar. Penelitian menggunakan metode regresi linear berganda dalam memprediksi saldo produksi hasil ternak. Tahapan penelitian yang dilakukan adalah mengumpulkan data hasil ternak, melakukan *pre-processing* data, memproses prediksi, membangun model dan evaluasi model menggunakan akurasi dengan *Mean Absolute Error* (MAE) dan R2-Score. Data yang digunakan mempunyai 14 atribut (kolom). Menggunakan korelasi, dipilih 3 atribut yang akan digunakan dalam membangun model yaitu komoditas, nilai produksi, dan biaya produksi (Rp.) Keseluruhan data dibagi menjadi dua yaitu data training dan data testing dengan proporsi berurutan sebesar 75%:25%, 80%:20%, dan 85%:15%. Model terbaik dihasilkan oleh pembagian *data training* dan *data testing* sebesar 85%:15%. Model yang dihasilkan adalah $\hat{Y} = -9810.74 + 22.58X_1 + X_2 - 0.99X_3$ dengan skor MAE sebesar 9232.04 dan skor R2-Score sebesar 1.

Kata Kunci : Regresi Linier Berganda, MAE, R2-Score, Hasil Ternak, Data Sains

PENDAHULUAN

Kota Blitar yang terletak di Provinsi Jawa Timur, merupakan wilayah terkecil kedua setelah Kota Mojokerto. Letak Kota Blitar yang berada di kaki Gunung Kelud menyebabkan Kota Blitar memiliki cuaca yang sejuk dengan rata-rata suhu 24-34C [1]. Walaupun berstatus Pemerintah Kota, Kota Blitar tidak serta merta mencerminkan kehidupan perkotaan karena luas wilayah dan kehidupan masyarakatnya yang masih mencerminkan kota kecil.

Lokasi Kota Blitar yang terletak dekat pantai selatan yang jauh dari area industri membuat sebagian besar masyarakatnya bergerak di berbagai sektor yang mengandalkan sumber daya manusia dan buatan [2]. Salah satunya adalah peternakan. Kota Blitar sendiri dikenal sebagai mutiaranya peternakan di Indonesia.

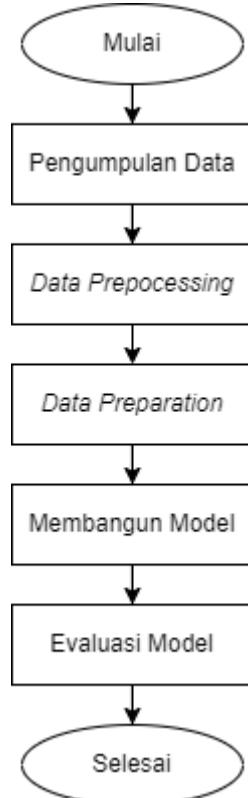
Seiring perkembangan teknologi, Indonesia telah memasuki tahap industri 4.0, dimana kecepatan dan ketepatan menjadi kunci sukses memenangkan persaingan yang semakin ketat. Sebagai sentra peternakan Indonesia, Kota Blitar harus dapat beradaptasi dengan perkembangan industri terkini agar tetap bisa unggul dalam persaingan terbuka saat ini. Melek teknologi, pola kemitraan, peningkatan kualitas produk dan sumber daya manusia adalah cara untuk menghadapi era industri 4.0 [3].

Sampai saat ini, terdapat banyak usaha mikro kecil menengah (UMKM) yang bergerak di sektor peternakan. Bahkan beberapa diantaranya telah menerapkan teknologi pada usaha mereka, antara lain untuk penjualan pakan ternak, inventaris toko, serta penjualan dan distribusi telur [3,4]. Namun keterbukaan informasi saat ini membuat harga

jual berbagai produk, termasuk produk di sektor peternakan, menjadi sangat fluktuatif [3]. Ketepatan dan kecepatan informasi terkait harga jual produk peternakan tentunya akan membuat hasil produksi menjadi optimal sehingga meningkatkan kualitas dan kuantitas sektor peternakan di Kota Blitar. Untuk itu, diperlukan suatu sistem berbasis teknologi kecerdasan artifisial yang tidak hanya mampu memberikan informasi terkait penjualan produk ternak, tetapi juga mampu memprediksi saldo produksi (dalam rupiah) yang dihasilkan. Hasil yang diharapkan dengan adanya prediksi ini masing – masing kecamatan dapat mengambil keputusan dalam menentukan komoditas hasil ternak mana yang menghasilkan keuntungan yang maksimal

METODE PENELITIAN

Dalam penelitian ini, terdapat beberapa tahapan yang dilaksanakan antara lain pengumpulan data, *data preprocessing*, *data preparation*, membangun model, evaluasi model. Diagram alir penelitian dapat dilihat pada Gambar 1.



GAMBAR 1. Diagram Alir Penelitian

Data penelitian ini diambil dari portal data Indonesia yaitu www.data.go.id. Data.go.id merupakan portal resmi terbuka data permerintah Indonesia yang disediakan oleh Kementerian Badan Perencanaan Pembangunan Nasional (Bappenas).

Pengumpulan Data

Data yang digunakan adalah hasil ternak Kabupaten Blitar dari tahun 2014 sampai tahun 2021. Atribut pada data terdiri dari; (1) provinsi, (2) kabupaten kota, (3) kecamatan, (4) status, (5) kode pum, (6) desa kelurahan, (7) tanggal, (8) jumlah ternak (ekor), (9) komoditas, (10) hasil produksi, (11) satuan, (12) nilai produksi, (13) biaya produksi (Rp.), dan (14) saldo produksi (Rp).

Data Preprocessing

Pada proses ini, dicek kembali tiap elemen pada semua kolom sudah sesuai format (tipe data) atau belum. Selanjutnya, dilakukan juga pengisian data atau penghapusan pada kolom-kolom yang masih kosong. Tahapan ini bertujuan untuk mengubah data mentah (*raw data*) yang tidak layak dianalisis menjadi data bersih (*clean data*) yang siap untuk dianalisis.

Data Preparation

Pada proses ini, setiap elemen data yang masih bertipe kategori diubah menjadi numerik. Selain itu, dilakukan proses korelasi untuk menyeleksi kolom mana yang akan digunakan. Korelasi yang digunakan dalam proses ini adalah korelasi Pearson. Persamaan korelasi Pearson dapat dilihat pada Persamaan (1).

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left(\left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) \left(\sum_{i=1}^n (Y_i - \bar{Y})^2 \right) \right)^{\frac{1}{2}}} \quad (1)$$

di mana r adalah koefisien korelasi, X_i dan Y_i berturut-turut adalah nilai ke- i pada kolom X dan Y , \bar{X} dan \bar{Y} nilai rata-rata kolom X dan Y , n adalah banyaknya data. Nilai dari korelasi Pearson adalah $-1 \leq r \leq 1$ [4]. Nilai positif adalah nilai 1, nilai -1 merupakan nilai negatif, dan nilai 0 merupakan nilai yang tidak terdapat korelasi [5]. Tahap ini bertujuan untuk mempersiapkan data sebelum digunakan untuk pembuatan model.

Membangun Model

Model yang digunakan pada penelitian ini adalah model regresi linier berganda. Diagram alir pembangunan model untuk prediksi saldo produksi hasil ternak Kabupaten Blitar dapat dilihat pada Gambar 2.



GAMBAR 2. Diagram Membangun Model Regresi Linier Berganda

Regresi linier berganda adalah suatu model statistik yang digunakan untuk meneliti hubungan antara satu variabel dependen (Y) dengan beberapa variabel independen (X_1, X_2, \dots, X_k) [6]. Model regresi linier berganda menggunakan

nilai-nilai variabel independen untuk meramalkan nilai variabel dependen. Persamaan umum regresi linier berganda dapat dilihat pada Persamaan (2).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i \quad (2)$$

di mana Y adalah variabel dependen, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ adalah parameter regresi, X_1, X_2, \dots, X_k adalah variabel independen dan ε_i adalah error pengamatan ke- i .

Data training dan *data testing* masing-masing dibagi menjadi data *input* dan data *output*. Data *input* terdiri dari atribut-atribut terbaik dari data hasil ternak Kabupaten Blitar. Data *output* merupakan atribut saldo produksi data hasil ternak Kabupaten Blitar. Pembagian *data training* dan *data testing* berturut-turut 75%:25%, 80%:20% dan 85%:15%. *Data training* digunakan untuk menghasilkan parameter regresi untuk membangun model. Selanjutnya, model yang dihasilkan di evaluasi menggunakan *data testing*.

Evaluasi Model

Evaluasi model dilakukan dengan cara mengecek tingkat akurasi yang dihasilkan model terhadap *data testing*. Pada penelitian ini digunakan dua metode untuk mengecek akurasi yaitu *Mean Absolute Error* (MAE) dan *R2-Score*. MAE didefinisikan sebagai rata-rata dari jumlah nilai mutlak data sebenarnya dikurangi data prediksi. MAE dapat dihitung menggunakan Persamaan (3).

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad (3)$$

di mana n adalah banyaknya data, Y_i adalah data sebenarnya ke- i , dan \hat{Y}_i adalah data prediksi ke- i . Ketika nilai MAE mendekati 0 maka model mempunyai akurasi yang tinggi dalam memprediksi data.

Koefisien determinasi atau *R2-Score* adalah proporsi variasi yang dijelaskan oleh variabel independen. *R2-Score* dapat dihitung menggunakan Persamaan (4).

$$R2 - Score = 1 - \frac{SS_R}{SS_T} \quad (4)$$

di mana SS_R adalah jumlah residual kuadrat dan SS_T adalah jumlah kuadrat. SS_R dan SS_T dapat dihitung menggunakan Persamaan (5) dan Persamaan (6).

$$SS_R = \sum_i (Y_i - \hat{Y}_i)^2 \quad (5)$$

$$SS_T = \sum_i (Y_i - \bar{Y})^2 \quad (6)$$

dengan \bar{Y} adalah rata-rata dari data sebenarnya. *R2-Score* menunjukkan varian residu (kesalahan prediksi) dalam memprediksi data dibagi dengan varian titik-titik data dari rata-rata semua data. *R2-Score* menunjukkan performa model dibandingkan dengan prediksi tebakan acak yang sama dengan rata-rata semua data. Semakin dekat *R2-Score* dengan 1 maka prediksi akan mendekati data sebenarnya.

HASIL DAN PEMBAHASAN

Data hasil ternak Kabupaten Blitar yang diambil pada website www.data.go.id mempunyai 407 baris. Sebelum digunakan untuk membangun model, data hasil ternak Kabupaten Blitar diberihsikan terlebih dahulu. Data hasil ternak Kabupaten Blitar mempunyai 14 atribut dengan masing-masing tipe data tiap atribut dapat dilihat pada Tabel 1.

TABEL 1. Tipe Data Tiap Atribut Data Hasil Ternak Kabupaten Blitar

| Atribut (Kolom) | Tipe Data |
|-----------------|----------------------------|
| Provinsi | Kategori (<i>Object</i>) |

| Atribut (Kolom) | Tipe Data |
|----------------------|----------------------------|
| Kabupaten/Kota | Kategori (<i>Object</i>) |
| Kecamatan | Kategori (<i>Object</i>) |
| Status | Kategori (<i>Object</i>) |
| Kode PUM | Numerik |
| Desa/Kelurahan | Kategori (<i>Object</i>) |
| Tanggal | Kategori (<i>Object</i>) |
| Jumlah Ternak (Ekor) | Numerik |
| Komoditas | Kategori (<i>Object</i>) |
| Hasil Produksi | Kategori (<i>Object</i>) |
| Satuan | Kategori (<i>Object</i>) |
| Nilai Produksi | Kategori (<i>Object</i>) |
| Biaya Produksi (Rp.) | Kategori (<i>Object</i>) |
| Saldo Produksi (Rp.) | Kategori (<i>Object</i>) |

Dapat dilihat pada Tabel 1, atribut nilai produksi, biaya produksi (Rp.), dan saldo produksi (Rp.) masih bertipe kategori (*object*). Seharusnya atribut tersebut bertipe numerik. Sehingga ketiga atribut tersebut diubah tipe datanya menjadi numerik. Selanjutnya, dicek elemen kosong pada masing-masing atribut yang dapat dilihat pada Tabel 2.

TABEL 2. Jumlah Elemen Kosong Data Hasil Ternak Kabupaten Blitar

| Atribut (Kolom) | Jumlah Elemen Kosong |
|----------------------|----------------------|
| Provinsi | 0 |
| Kabupaten/Kota | 0 |
| Kecamatan | 0 |
| Status | 0 |
| Kode PUM | 0 |
| Desa/Kelurahan | 0 |
| Tanggal | 0 |
| Jumlah Ternak (Ekor) | 0 |
| Komoditas | 14 |
| Hasil Produksi | 0 |
| Satuan | 399 |
| Nilai Produksi | 0 |
| Biaya Produksi (Rp.) | 0 |
| Saldo Produksi (Rp.) | 0 |

Berdasarkan Tabel 2, atribut yang mempunyai elemen kosong adalah komoditas dan satuan. Untuk atribut satuan dihapus, dikarenakan mempunyai banyak elemen yang kosong. Sedangkan, untuk atribut komoditas dilakukan interpolasi untuk mengisi elemen yang kosong. Atribut provinsi, kabupaten/kota dan tanggal dihapus karena tidak digunakan dalam proses perhitungan. Hasil pembersihan data dapat dilihat pada Tabel 3.

TABEL 3. Data Hasil Pembersihan

| Atribut (Kolom) | Tipe Data | Jumlah Elemen Kosong |
|----------------------|----------------------------|----------------------|
| Kecamatan | Kategori (<i>Object</i>) | 0 |
| Status | Kategori (<i>Object</i>) | 0 |
| Kode PUM | Numerik | 0 |
| Desa/Kelurahan | Kategori (<i>Object</i>) | 0 |
| Jumlah Ternak (Ekor) | Numerik | 0 |
| Komoditas | Kategori (<i>Object</i>) | 0 |
| Hasil Produksi | Numerik | 0 |
| Nilai Produksi | Numerik | 0 |
| Biaya Produksi (Rp.) | Numerik | 0 |
| Saldo Produksi (Rp.) | Numerik | 0 |

TABEL 4. Pelabelan Atribut Kecamatan

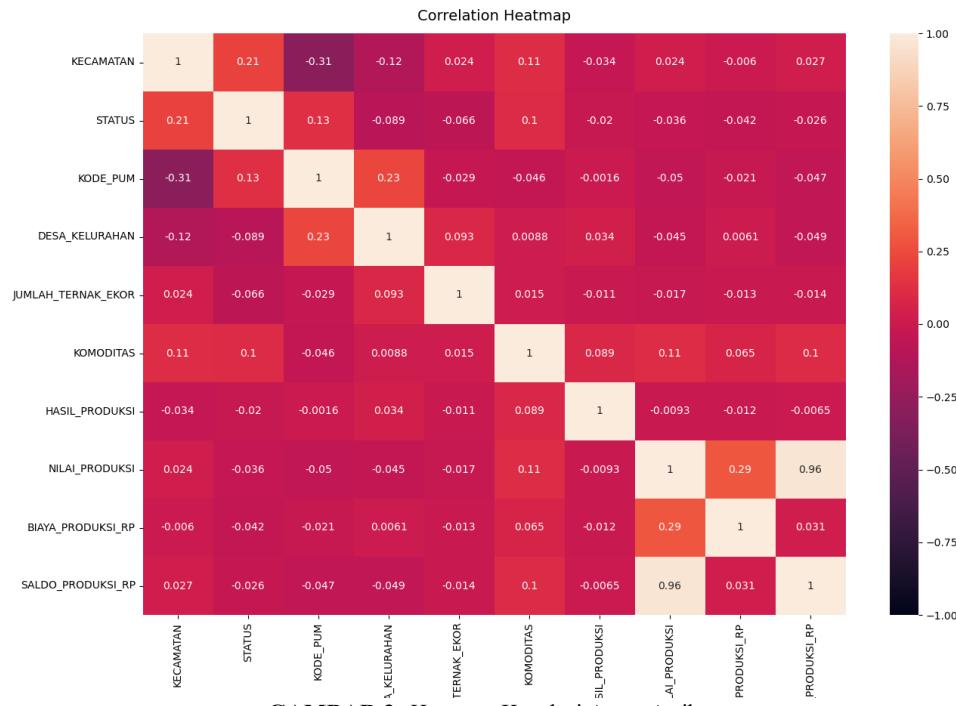
| Atribut (Kolom) | Pelabelan |
|-----------------|-----------|
| Bakung | 0 |
| Binangun | 1 |
| Doko | 2 |
| Gandusari | 3 |
| Garum | 4 |
| Kademangan | 5 |
| Kanigoro | 6 |
| Kesamben | 7 |
| Nglegok | 8 |
| Panggung Rejo | 9 |
| Ponggok | 10 |
| Sanankulon | 11 |
| Selopuro | 12 |
| Selorejo | 13 |
| Srengat | 14 |
| Sutojayan | 15 |
| Talun | 16 |
| Udanawu | 17 |
| Wates | 18 |
| Wlingi | 19 |
| Wonodadi | 20 |
| Wonotirto | 21 |

TABEL 5. Pelabelan Atribut Status

| Atribut (Kolom) | Pelabelan |
|-----------------|-----------|
| Desa | 0 |
| Kelurahan | 1 |

TABEL 6. Pelabelan Atribut Komoditas

| Atribut (Kolom) | Pelabelan | | | | | | | | | | |
|-----------------------|-----------|--|--|--|--|--|--|--|--|--|--|
| Air Liur Burung Walet | 0 | | | | | | | | | | |
| Bulu | 1 | | | | | | | | | | |
| Burung Walet | 2 | | | | | | | | | | |
| Cinderamata | 3 | | | | | | | | | | |
| Daging | 4 | | | | | | | | | | |
| Hiasan/Lukisan | 5 | | | | | | | | | | |
| Kerupuk Kulit | 6 | | | | | | | | | | |
| Kulit | 7 | | | | | | | | | | |
| Madu | 8 | | | | | | | | | | |
| Minyak | 9 | | | | | | | | | | |
| Susu | 10 | | | | | | | | | | |
| Telur | 11 | | | | | | | | | | |



GAMBAR 3. Heatmap Korelasi Antar Atribut

Data hasil pembersihan, diolah lagi untuk disiapkan dalam membangun model. Atribut-atribut yang bertipe kategori pada Tabel 3, dilabelkan agar bisa dihitung korelasinya dengan atribut saldo produksi (Rp.). Hasil pelabelan atribut yang bertipe kategori dapat dilihat pada Tabel 4, Tabel 5, dan Tabel 6.

Untuk atribut desa/kelurahan dilabeli dari 0 sampai 146 karena terdapat 147 desa/kelurahan di data hasil ternak Kabupaten Blitar. Selanjutnya dihitung korelasi untuk melihat hubungan antar atribut. Hasil perhitungan korelasi dapat dilihat pada Gambar 3. Berdasarkan Gambar 3, tiga atribut yang mempunyai korelasi paling tinggi dengan atribut saldo produksi (Rp.) antara lain komoditas, nilai produksi, dan biaya produksi (Rp.). Nilai korelasi atribut-atribut tersebut berturut-turut sebesar 0.1, 0.96, dan 0.031.

Atribut komoditas, nilai produksi, dan biaya produksi (Rp.) digunakan sebagai *data training* dan *data testing* untuk membangun model dan mengevaluasi model prediksi atribut saldo produksi (Rp.). Atribut komoditas sebagai X_1 , atribut nilai produksi sebagai X_2 , atribut biaya produksi (Rp.) sebagai X_3 , dan atribut saldo produksi (Rp.) sebagai Y . Perbandingan pembagian data training dan data testing adalah 75%:25%, 80%:20%, dan 85%:15%. Koefisien model regresi linier berganda yang dihasilkan menggunakan *data training* dapat dilihat pada Tabel 7. Masing-masing model yang dihasilkan dari *data training*, dicek akurasi MAE dan *R2-Score* dari menggunakan *data testing* dapat dilihat pada Tabel 8.

TABEL 7. Koefisien Regresi Linier Berganda

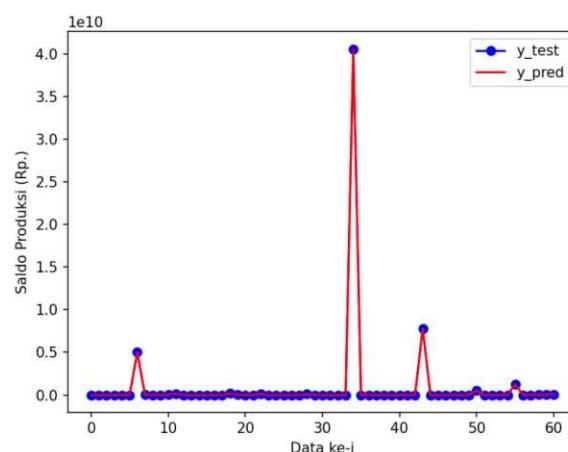
| Data Training | β_0 | β_1 | β_2 | β_3 |
|----------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| 75% | -11281.80 | 108.50 | 1 | -0.99 |
| 80% | -10387.73 | 8.56 | 1 | -0.99 |
| 85% | -9810.74 | 22.58 | 1 | -0.99 |

TABEL 8. Hasil Akurasi MAE dan *R2-Score*

| Data Testing | MAE | R2-Score |
|---------------------|------------|-----------------|
| 25% | 11972.28 | 1 |
| 20% | 10531.38 | 1 |
| 15% | 9232.04 | 1 |

Berdasarkan Tabel 8, pembagian data testing sebesar 15% mempunyai nilai MAE yang paling kecil yaitu 9232.04 sedangkan, untuk nilai *R2-Score* semua bernilai sama untuk masing-masing *data testing*. Sehingga, diambil koefisien regresi linier berganda dari pembagian *data training* sebesar 85%. Oleh karena itu, diperoleh model regresi linier berganda untuk memprediksi saldo produksi hasil ternak Kabupaten Blitar sesuai Persamaan (7). Dapat dilihat pada Gambar 4 di mana grafik data prediksi sama dengan grafik data aktual. Hal ini menunjukkan bahwa model dapat memprediksi dengan tepat sesuai dengan keadaan aktual. Gambar 4 juga mendukung hasil akurasi yang didapatkan.

$$\hat{Y} = -9810.74 + 22.58X_1 + X_2 - 0.99X_3 \quad (7)$$



GAMBAR 4. Perbandingan Data Aktual Dan Data Prediksi

KESIMPULAN

Berdasarkan hasil penelitian, maka diperoleh beberapa kesimpulan sebagai berikut: 1) Menggunakan korelasi, dipilih tiga atribut dari 14 atribut yang digunakan untuk membangun model yaitu komoditas, nilai produksi, dan biaya produksi (Rp.). 2) Koefisien model regresi linier berganda diperoleh menggunakan data training sebesar 75%, 80%, dan 85%. Akurasi MAE dan R2-Score dari masing-masing model diuji menggunakan data testing sebesar 25%, 80%, dan 15%. 3) Koefisien Model regresi linier berganda dipilih berdasarkan nilai akurasi terbaik yaitu 85%:15% dengan akurasi MAE dan R2-Score sebesar 9232.04 dan 1. 4) Model regresi linier berganda untuk memprediksi saldo produksi hasil ternak Kabupaten Blitar adalah $\hat{Y} = -9810.74 + 22.58X_1 + X_2 - 0.99X_3$. 5) Model regresi linier berganda dapat memprediksi dengan akurat karena data prediksi sesuai dengan data aktual.

REFERENSI

- [1] Triyanto, E., Sismoro, H., & Laksito, A.D. 2019. IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI PRODUKSI PADI DI KABUPATEN BANTUL. *Jurnal Teknologi dan Sistem Informasi Univrab*.
- [2] Romandoni, A. 2017. Penerapan Metode Regresi Linear Berganda untuk Prediksi Hasil Panen Jagung. *STM IK Sinar Nusantara*.
- [3] Utomo, N. S. 2015. Aplikasi Prediksi Kelayakan Operasional Mesin Rivet Produksi Menggunakan Metode Regresi Linear Berganda (Studi Kasus: PT Pradana Indah Sejahtera). *Politeknik Negeri Batam*.
- [4] Zhang, Y., Li, Y., Song, J., Chen, X., Lu, Y., & Wang, W. (2020). Pearson correlation coefficient of current derivatives based pilot protection scheme for long-distance LCCHVDC transmission lines. *International Journal of Electrical Power and Energy Systems*, 116(November 2022), 105526. <https://doi.org/10.1016/j.ijepes.2019.105526>.
- [5] Fu, T., Tang, X., Cai, Z., Zuo, Y., Tang, Y., & Zhao, X. (2020). Correlation research of phase angle variation and coating performance by means of Pearson's correlation coefficient. *Progress in Organic Coatings*, 139(November 2022), 105459. <https://doi.org/10.1016/j.porgcoat.2019.105459>.
- [6] Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. Fifth edition, Prentice Hall Inc. Upper Saddle River : NJ.
- [7] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2012, vol. 821.