

**Application Of Data Mining For Classification Of Sales Best-Selling Products
Using The C5.0 Algorithm
(Case Study of Syarah Bakery in the West Circle)**

**Penerapan Data Mining Untuk Klasifikasi Penjualan Produk Terlaris
Menggunakan Metode Algoritma C5.0
(Studi Kasus Syarah Bakery Lingkar Barat)**

Bagus Hakim Perkoso ¹⁾; Asnawati ²⁾; Lena Elfianty ³⁾

¹⁾Study Program of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu

^{2,3)} Department of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu

Email: ¹⁾ bagushakim2000@gmail.com

ARTICLE HISTORY

Received [01 Mei 2023]

Revised [07 Juni 2023]

Accepted [28 Juni 2023]

KEYWORDS

Data Mining, Classification,
Algorithm C5.0

This is an open access article under
the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Tujuan dari Penelitian ini Untuk menerapkan Data mining dalam mengklasifikasikan penjualan produk terlaris menggunakan Algoritma C5.0 Metode Penelitian yang digunakan adalah Metode Waterfall yang bersifat serial yang dimulai dari proses perencanaan, analisa, desain, dan implementasi pada sistem. Pada algoritma C5.0 dapat menangani atribut kontiyu dan diskrit. Pertama yang dilakukan adalah menghitung nilai entropy dari keseluruhan atribut, lalu menghitung nilai information gain tertinggi dari seluruh atribut sehingga didapatkan atribut yang akan digunakan sebagai akar atau parent. Saran dari penulis adalah agar dapat difungsikan Aplikasi Data Mining Nantinya untuk Mengetahui produk yang diminati dan membantu perencanaan penyediaan stok kedepannya

ABSTRACT

Data mining is the process of obtaining useful information from a large database and needs to be extracted so that it becomes new information and can assist in decision making. The purpose of this study is to apply data mining in classifying best-selling product sales using the C5.0 algorithm. The research method used is the Waterfall Method which is serial in nature which starts from the process of planning, analysis, design and implementation of the system. The C5.0 algorithm can handle continuous and discrete attributes. The first thing to do is to calculate the entropy value of all attributes, then calculate the highest information gain value of all attributes so that the attribute that will be used as the root or parent is obtained. The suggestion from the author is that the Data Mining Application can function later to find out which products are of interest and help plan future supply of stock.

PENDAHULUAN

Teknologi berubah dengan cepat dan membantu orang mengakses informasi dengan berbagai cara. Orang-orang dari segala usia dan terutama anak-anak menggunakan teknologi setiap hari. Sebagian besar orang Indonesia percaya bahwa teknologi dapat menggantikan upaya manusia dan membantu mereka dengan tugas apa pun. Informasi perlu disajikan dengan cepat dan efisien saat ini. Karena metode pencatatan informasi telah berubah dari waktu ke waktu, sistem komputerisasi modern seharusnya diterapkan sebagai gantinya.

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Klasifikasi data mining menjadi salah satu metode yang paling umum

untuk digunakan. Metode ini dilakukan bertujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui.

Syarah Bakery Jl. Mahakam IV Kelurahan Jl. Gedang, Kecamatan Gading Cempaka, Kota Bengkulu. Adalah Usaha Mikro Kecil Menengah (UMKM) yang menawarkan beberapa macam produk seperti , Snack manis segar, Snack asin, Manis Tradisional, Pastry dan Cake manis. Ketersediaan produk menjadi tantangan tersendiri untuk menjaga kestabilan stok produk. Walaupun kelihatannya mudah, namun mengelola persediaan produk bukanlah sesuatu yang bisa dianggap remeh, karena jika persediaan terlalu banyak, risiko kerusakan produk akan meningkat, dan sebaliknya, jika persediaan sedikit, risiko kekurangan produk akan lebih tinggi dari persediaan dan dapat menunda keuntungan serta membuat konsumen kecewa.

Dalam hal ini Data Mining sangat berpengaruh dalam proses menggali sebuah data dari suatu kumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual. Pada umumnya Data mining muncul dari banyaknya jumlah data yang tersimpan dalam suatu database. Data mining adalah pembelajaran mesin, pengenalan pola, database, statistik, dan teknik visualisasi yang digunakan untuk memecahkan masalah penggalian informasi dari repositori database besar. Data Mining adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari database yang besar.

Algoritma C5.0 Data mining adalah salah satu algoritma klasifikasi yang secara khusus dapat diterapkan pada teknisi Decision Tree. Algoritma C5.0 merupakan perbaikan dari algoritma ID3 dan C4.5 yang diperkenalkan terlebih dahulu oleh j.Ross Quinlan pada tahun 1987. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan Information Gain. Memilih atribut untuk pemecah objek dalam beberapa kelas harus dipilih dari atribut yang menghasilkan Information Gain paling besar. Atribut dengan nilai Information Gain tertinggi akan dipilih sebagai parent untuk node selanjutnya.

LANDASAN TEORI

Data

Menurut Rusmawan (2019, hlm. 35) "Data adalah catatan atas kumpulan fakta". Menurut Andri Kristanto(2018:7), "Data adalah penggambaran dari sesuatu dan kejadian yang kita hadapi, kenyataan yang menggambarkan suatu kejadian-kejadian dan kesatuan nyata".

Basis data

Menurut Abdulloh (2018:103), "Database atau basis data adalah kumpulan informasi yang disimpan dalam komputer secara sistematis sehingga dapat diperiksa menggunakan suatu program komputer untuk memperoleh informasi".

Menurut Enterprise (2017:1), "Database atau basis data adalah suatu aplikasi yang menyimpan sekumpulan data".

Menurut A.S dan Shalahudin (2018:28) "basis data adalah sistem terkomputerisasi yang tujuan utamanya adalah memelihara data yang sudah diolah atau informasi dan membuat informasi tersedia saat dibutuhkan. Pada intinya basis data adalah media untuk menyimpan data agar dapat diakses dengan mudah dan cepat."

Data Mining

Menurut (Firdaus,2017: 91-97) "Data mining adalah pencarian pengetahuan dalam basis data pada proses identifikasi pola-pola yang valid, yang berpotensi bermanfaat dan dapat dipahami dengan mudah. Data mining merupakan proses yang bertujuan untuk menemukan pengetahuan (knowledge discovery) yang ditambang dari sekumpulan data yang memiliki volume yang sangat besar".

Menurut (Suntoro, 2019:2) "Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan".

Algoritma C 5.0

Algoritma C5.0 adalah salah satu algoritma yang merupakan penyempurnaan dari algoritma C4.5 yang menggunakan representasi berbentuk pohon yang setiap node mempresentasikan atribut kemudian cabang mempresentasikan nilai dari atribut dan memiliki yang namanya daun dimana fungsinya adalah kelas. Pengambilan keputusan didasarkan pada nilai Gain terbesar dari hasil perhitungan semua atribut . Algoritma C5.0 menghasilkan tingkat keakuratan yang lebih tinggi dalam hal prediksi penggunaan

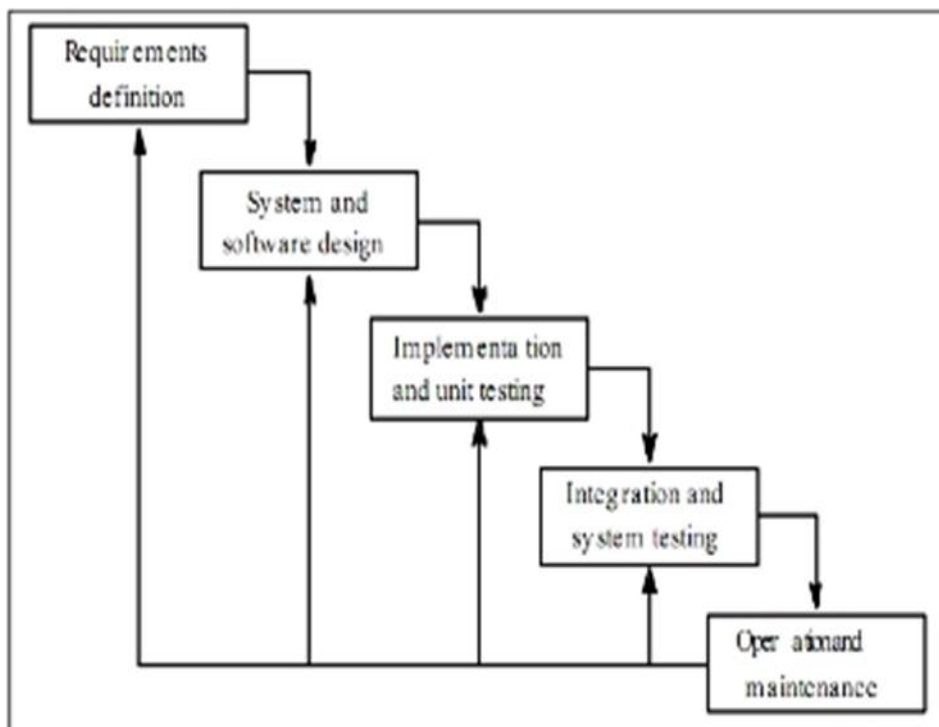
algoritma C5.0 dapat menghasilkan model prediksi dengan hasil tingkat akurasi yang lebih tinggi. Algoritma C5.0 diharapkan proses penggalian informasi lebih cepat dan optimal dengan kapasitas data yang lebih besar, sehingga kesalahan yang ditimbulkan dalam pengambilan keputusan lebih diminimalkan. Pada algoritma C5.0 dapat menangani atribut kontinyu dan diskrit. Pertama yang dilakukan adalah menghitung nilai entropy dari keseluruhan atribut, lalu selanjutnya yaitu menghitung nilai information gain tertinggi dari seluruh atribut sehingga didapatkan atribut yang akan digunakan sebagai akar atau parent. Selanjutnya percabangan pada akar untuk setiap nilainya ditentukan, kemudian setiap cabang berisi kasus yang telah dibagi. Kemudian perhitungan secara berulang dilakukan untuk menentukan nilai gain. perhitungan tersebut berhenti ketika semua data telah dihitung memiliki persamaan pada kelasnya. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan Information Gain. Memilih atribut untuk pemecah objek dalam beberapa kelas harus dipilih dari atribut yang menghasilkan Information Gain paling besar. Atribut dengan nilai Information Gain tertinggi akan dipilih sebagai parent untuk node selanjutnya.

METODE PENELITIAN

Penelitian ini menggunakan metode waterfall. Metode Waterfall merupakan metode pengembangan perangkat lunak tertua sebab sifatnya yang natural. Metode Waterfall merupakan pendekatan SDLC paling awal yang digunakan untuk pengembangan perangkat lunak. Urutan dalam Metode Waterfall bersifat serial yang dimulai dari proses perencanaan, analisa, desain, dan implementasi pada sistem.

Metode ini dilakukan dengan pendekatan yang sistematis, mulai dari tahap kebutuhan sistem lalu menuju ke tahap analisis, desain, coding, testing/verification, dan maintenance. Langkah demi langkah yang dilalui harus diselesaikan satu per satu (tidak dapat meloncat ke tahap berikutnya) dan berjalan secara berurutan, oleh karena itu disebut waterfall (AirqazqazTerjun).

Gambar 1 Tahapan Metode Waterfall



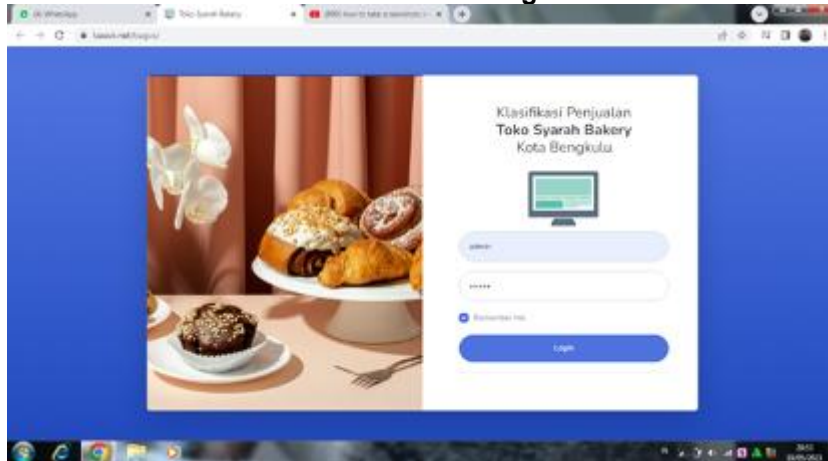
HASIL DAN PEMBAHASAN

Hasil

Berikut ini merupakan tampilan hasil dan pembahasan dari aplikasi Klasifikasi Penjualan Produk Terlaris Menggunakan Algoritma C5.0 Tampilan ini dibuat sedemikian rupa untuk mendukung segala sesuatu yang dibutuhkan oleh perangkat lunak ini agar berjalan dengan baik.

1. Tampilan Hasil Form Login

Gambar 2 Form Login

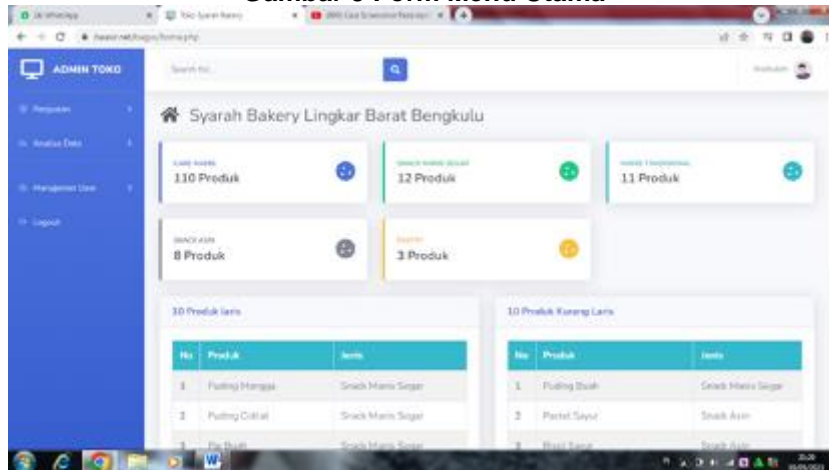


Form login ini berfungsi sebagai keamanan sistem. Untuk melakukan login, isi username dan password lalu klik tombol login, jika berhasil maka form menu utama akan muncul. tetapi jika username dan password yang di inputkan salah, maka akan muncul silahkan untuk menginputkan ulang lagi.

2. Tampilan Form Menu Utama

Berikut merupakan tampilan form menu utama dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu.

Gambar 3 Form Menu Utama

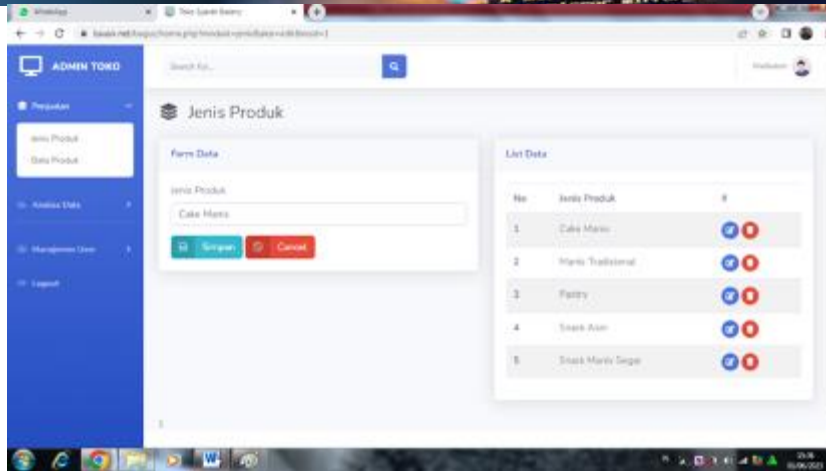


Form menu utama terdapat 4 pilihan, yaitu Penjualan, analisa Data, Manajemen User dan Logout. Pada menu penjualan terdapat 2 pilihan yaitu jenis produk dan data produk. dan pada menu analisa Data terdapat 3 pilihan, yaitu Pohon keputusan, Form Data, dan Analisa Data. kemudian pada menu Manajemen User Terdapat 1 pilihan saja, yaitu Update data. dan logout untuk menu keluar aplikasi.

3. Tampilan Form Jenis Produk

Berikut merupakan tampilan form jenis produk dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu.

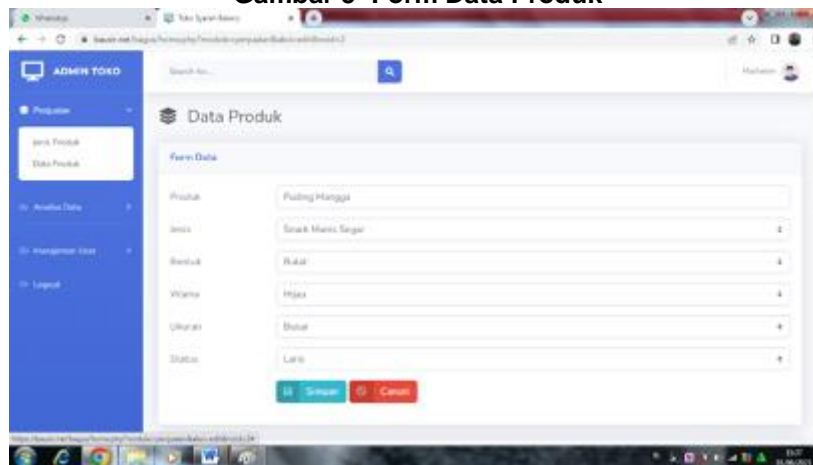
Gambar 4 Form Jenis Produk



4. Tampilan Form Data Produk

Berikut merupakan tampilan form menu Data Produk dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu. Terdapat 144 Data penjualan produk.

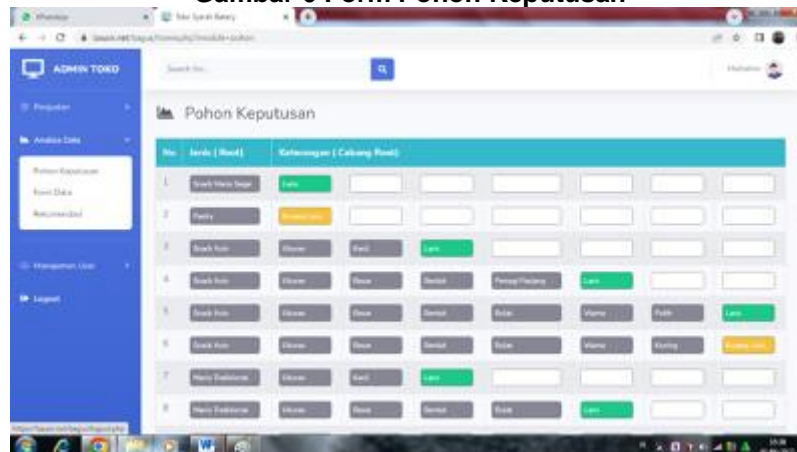
Gambar 5 Form Data Produk



5. Tampilan Form Pohon Keputusan

Berikut merupakan tampilan form pohon keputusan dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu. Form pohon keputusan tersebut merupakan output atau hasil analisa data akhir dari klasifikasi Algoritma C5.0 dan telah didapatkan jenis (Root) dan Keterangan (Cabang Root).

Gambar 6 Form Pohon Keputusan



6. Tampilan Form Data

Berikut merupakan tampilan form Data dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu. Form Data tersebut untuk dijadikan bahan uji pohon keputusan yang sudah berjalan, jadi isi dari prediksi itu tersebut mengikuti pohon keputusan yang telah dibuat dan untuk dijadikan bahan kombinasi produk antara laris dan kurang Laris.

Gambar 7 Form Data

7. Tampilan Rekomendasi Produk

Berikut merupakan tampilan form Rekomendasi Produk dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu. Rekomendasi produk tersebut merupakan hasil akhir dari pohon keputusan untuk membenarkan bahwasannya telah berjalan dan dibuatlah 2 tabel keterangan list produk laris dan Kurang laris.

Gambar 8 Form Rekomendasi Produk

No	Jenis	Ukuran	Warna	Warna
1	Snack Manis Seder	Sedang	Sedang	Sedang
2	Snack Asin	Kecil	Sedang	Sedang
3	Snack Asin	Besar	Persegi Panjang	Sedang
4	Snack Asin	Besar	Bulat	Putih
5	Snack Tradisional	Kecil	Sedang	Sedang
6	Snack Tradisional	Besar	Bulat	Sedang
7	Snack Tradisional	Besar	Persegi Panjang	Putih

8. Tampilan Manajemen User

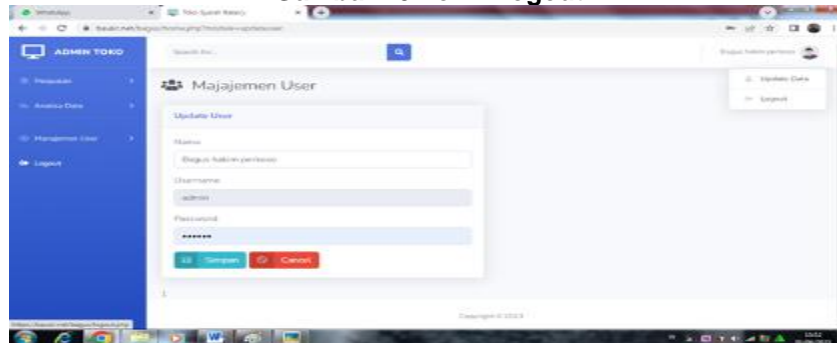
Berikut merupakan tampilan form manajemen user dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu. Tampilan manajemen user adalah form untuk mengupdate nama, dan password.

Gambar 9 Form Manajemen user

9. Tampilan Logout

Berikut merupakan tampilan form logout dalam perancangan Aplikasi Klasifikasi Toko Syarah Bakery Kota Bengkulu. Tampilan logout adalah untuk keluar dari aplikasi .

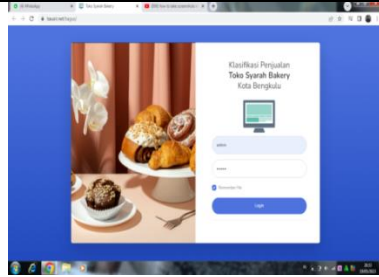
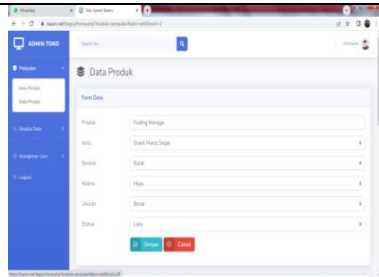
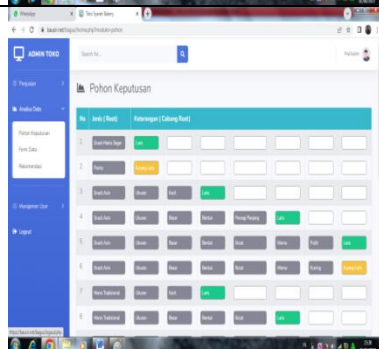
Gambar 10 Form Logout



Hasil Pengujian Sistem

Pengujian sistem menggunakan metode black-box dilakukan dengan cara memberikan sejumlah data masukan pada aplikasi, kemudian dilakukan proses sesuai dengan kebutuhan fungsionalnya untuk melihat apakah aplikasi yang sedang diproses sesuai dengan kebutuhan fungsionalnya dan juga aplikasi menghasilkan keluaran yang digunakan dan sesuai dengan fungsi dari program tersebut. Apabila data masukan yang diberikan proses menghasilkan keluaran yang sesuai, maka aplikasi yang dibuat telah benar.

Tabel 1 Hasil Pengujian Sistem

No	Nama Pengujian	Test	Hasil	Keterangan
1	Admin terlebih dahulu Masukan username dan password yang benar		Halaman login berhasil	Pengujian Berhasil
2	Melakukan penginputan data produk		Sesuai dengan aplikasi yang dirancang tombol input data produk berfungsi dengan baik	Pengujian Berhasil
3	Melakukan output hasil Akhir Pohon keputusan pada Penjualan produk menentukan laris dan kurang laris		Sesuai dengan aplikasi yang dirancang tombol output hasil Akhir pohon keputusan berfungsi dengan benar	Pengujian Berhasil

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil dan kesimpulan dari pembahasan dan pengujian yang telah dilakukan, maka dapat disimpulkan adalah sebagai berikut Penerapan data mining untuk klasifikasi penjualan produk terlaris menggunakan metode algoritma C5.0 (Studi Kasus Syarah Bakery Lingkar Barat) dibuat menggunakan perangkat lunak (software) yaitu bahasa pemrograman php dengan menggunakan database MySQL sebagai media pendukung. Berdasarkan data 144 data Penjualan Produk menggunakan Algoritma C5.0 untuk mengklasifikasikan penjualan produk terlaris di syarah bakery menggunakan variabel jenis barang, Bentuk,warna,dan Ukuran. Hasil akhir dari proses klasifikasi algoritma C5.0 adalah pohon keputusan dengan rule didalamnya. Kinerja algoritma C5.0 mendapatkan tingkat akurasi yang tinggi sebesar 91,719 % .

Saran

Sistem yang baru ini hendaknya dapat digunakan oleh pihak toko syarah bakery sebagaimana mestinya karena sistem baru ini dapat membantu dan mempermudah dalam proses perencanaan penyediaan stok produk kedepannya. Perlunya maintance untuk pemeliharaan program berdasarkan metode yang dikembangkan.

DAFTAR PUSTAKA

- Andri Kristanto, (2018:7) Kristanto, A. B. (2018). Tindak Tutur Asertif dalam Interaksi Jual Beli Buah di Pasar Tanjung dan Pemanfaatannya sebagai Materi Pembelajaran Teks Negosiasi di SMA (Doctoral dissertation, Fakultas Keguruan dan Ilmu Pendidikan).
- Abdullah, (2018:103) Implementasi Metode Analytical Hierarchy Process (Ahp) Untuk Evaluasi Kinerja Pegawai Puskesmas Tunas Harapan (Doctoral dissertation, Politeknik Negeri Sriwijaya).
- A.S dan Salahuddin (2018:28) APLIKASI PENGELOLAAN DATA UNTUK Mendukung PROSES PEMASARAN USAHA MIKRO KECIL DAN MENENGAH (UMKM) DI WILAYAH KELURAHAN PLAJU ULU PALEMBANG BERBASIS WEBSITE (Doctoral dissertation, Politeknik Negeri Sriwijaya).
- Enterprise, (2017:1) Implementasi Metode Analytical Hierarchy Process (Ahp) Untuk Evaluasi Kinerja Pegawai Puskesmas Tunas Harapan (Doctoral dissertation, Politeknik Negeri Sriwijaya).
- Firdaus, (2017: 91-97) Optimalisasi Pengelompokkan Tingkat Kesejahteraan Ekonomi Penduduk Dalam Program Pembinaan Masyarakat (Doctoral dissertation, UNIVERSITAS PUTRA INDONESIA YPTK PADANG).
- Rusmawan, (2019:35) APLIKASI PENGOLAHAN DATA HASIL PENDAPATAN PENJUALAN BARANG DAN UPAH JASA BERBASIS WEBSITE PADA PT. ADELINDO PERKASA (Doctoral dissertation, POLITEKNIK NEGERI SRIWIJAYA).
- Suntoro, (2019:2) Analisis Dataset COVID-19 menggunakan Algoritma KNN dan Random Forest. Jurnal STRATEGI-Jurnal Maranatha, 4(1), 58-69.
- Towa p. Dan J.N.B. Tairas (2022 : 531-548) persepsi Mahasiswa Program Studi S1 Ilmu Perpustakaan Fakultas Ilmu Budaya Universitas Diponegoro Terhadap Pelaksanaan Klasifikasi Koleksi di Perpustakaan SD Negeri Manyaran 01 Semarang." Anuva: Jurnal Kajian Budaya, Perpustakaan, dan Inf