IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK MENENTUKAN SASARAN PRIORITAS PEMBELI

STUDI KASUS : AGEN BUAH YUDI

PROPOSAL TUGAS AKHIR

Diajukan oleh :

Nanda Ghina Syawali

8020190189

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir

Sebagai Akhir Proses Studi Strata 1

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS DINAMIKA BANGSA

2022

PERNYATAAN HASIL EVALUASI

NAMA/NIM : Nanda Ghina Syawali / 8020190189

PRODI : ~~SI~~ / TI / ~~SK~~ *)

JUDUL :IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK MENENTUKAN SASARAN PRIORITAS PEMBELI (STUDI KASUS : AGEN BUAH YUDI)

1. Hasil Evaluasi : Disetujui / Disetujui dengan perbaik / Ditolak *)
   Catatan :
   Alasan Penolakan Tugas Akhir
   - ☐ Tugas akhir tidak relevan dengan Program Studi
   - ☐ Pernah ada topik sejenis
   - ☐ Metode utama telah banyak dipakai
   - ☐ Metode yang dipakai tidak jelas
   - ☐ Masalah terlalu sempit
   - ☐ _____

Mengetahui,

Ketua Program Studi

Beny, S.Kom, MSc

NIP. YDB: 07.84.055

*) Coret yang tidak perlu

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK MENENTUKAN SASARAN PRIORITAS PEMBELI (STUDI KASUS : AGEN BUAH YUDI)

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

a. Nama Lengkap : Nanda Ghina Syawali
b. NIM : 80201090189
c. Jenis Kelamin : Perempuan
d. Tempat/Tgl. Lahir : Jambi/27 Januari 2000
e. Alamat : Jl. Letkol A. Tarmizi Kadir RT.25 No. 01 Kel. Tambak Sari Kec. Jambi Selatan.
f. No. Telepon : 0821-7746-0642
g. Email : nandaghina8@gmail.com

1.1. LATAR BELAKANG

Data mining merupakan proses dengan kecerdasan buatan, teknik statistik, matematika, dan *machine learning* untuk mengindentifikasi dan mengekstrasi dan suatu informasi yang bermanfaat dan pengetahuan terkait dari berbagai database yang besar #document. *K-Means Clustering* merupakan salah satu algoritma dalam *data mining* yang bisa berfungsi untuk melakukan pengelompokan/*Clustering* suatu data. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam group yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *Clustering.* Metode *K-Means* ialah metode yang termasuk dalam algoritma *Clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik[1].

K-Means Clustering bukanlah hal yang baru, berdasarkan penelitian yang dilakukan "Haviluddin, Suryani Junita Patandianan et al[2] membahas tentang metode *K-Means* dalam mengkelompokkan rekomendasi tugas akhir. Dengan menggunakan metode *K-Means Clustering* mampu pengelompokan data rekomendasi tugas akhir menjadi 3 *cluster* berdasarkan MKW dengan nilai A, B, dan C, dimana C1 adalah sedikit, yang berhubungan dengan 1 MKW Pemrograman Visual; C2 adalah sedang, yang berhubungan dengan 6 MKW yaitu Basis Data, Sistem Keamanan Komputer, Pemrograman Berbasis Objek, Pemrograman Web, dan Struktur Data, sedangkan C3 adalah banyak, yang berkorelasi dengan 3 MKW yaitu Sistem Operasi, Sistem Keamanan Komputer, dan Kecerdasan Buatan. Perhitungan pengukuran hasil *cluster* menggunakan metode Sum of Squared Errors(SSE) dan evaluasi *cluster* menggunakan metode Silhouette Coefficient(SC), dimana, SSE 3 *cluster* bernilai 0.4506% dan 4 *cluster* bernilai 1.1072% telah didapatkan".

“Istiqomah Sumadikarta dan Evan Abeiza[3] membahas tentang metode *K-Means Clustering* untuk mengkelompokkan produk dan pelanggan potensial. Dengan menggunakan metode *K-Means Clustering* mampu pengelompokan data produk dan pelanggan menjadi 5 *Cluster*, C1 dengan jumlah *customer* yaitu 3, C2 dengan jumlah *customer* yaitu 19, C3 dengan jumlah *customer* yaitu 73, C4 dengan jumlah *customer* yaitu 245, dan C5 dengan jumlah *customer* yaitu 9”.

Pada penelitian ini penulis menggunakan *Data mining* metode *K-Means Clustering* untuk memntukan prioritas pembeli terhadap buah-buahan di Agen Buah Yudi. Karena “*Data mining* disebut sebagai *knowledge discovery in database* (KDD), yaitu kegiatan yang meliputi pengumpulan dan pemakaian data historis yang bertujuan menemukan keteaturan dan pola hubungan pada data set yang memiliki ukutan besar. *K-Means* adalah sebuah metode pengklasteran memakai konsep partitioning yang nantinya dalam prosesnya algoritma akan memisahkan data-data dalam beberapa *cluster*/kelompok berbeda. *Clustering* adalah sebuah teknik yang dipakai untuk memasukan data ke dalam sebuah kelompok atau grup yang memiliki kedekatan khusus pada masing-masing objek”[4].

Dari permasalahan di atas, penulis tertarik untuk melakukan penelitian berguna memberikan solusi bagi pemilik Agen Buah Yudi dalam menentukan prioritas pembeli. Penulis menungkan di dalam tugas akhir yang berjudul “**IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK MENENTUKAN SASARAN PRIORITAS PEMBELI (STUDI KASUS : AGEN BUAH YUDI)**

1.2. RUMUSAN MASALAH

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah :

1. "Bagaimana menerapkan *data mining* untuk menetukan prioritas pembeli buah-buahan di Agen Buah Yudi dengan cara *Clustering* data penjualan buah menggunakan Algoritma *K-Means*?".
2. "Bagaimana menganalisis dari hasil perhitungan *Clustering* data pembelian buah?".

1.3. BATASAN MASALAH

Pembatasan masalah yang digunakan dalam sebuah pembahasan bertujuan agar dalam pembahasannya lebih terarah dan sesusai dengan tujuan yang akan dicapai. Maka penulis membatasi permasalahan seperti berikut ini:

1. Penelitian ini difokuskan pada pembelian buah di Agen Buah Yudi.
2. Metode yang digunakan adalah *K-Means Clustering*.
3. Alat bantu analisis menggunakan Tools Excel, *Weka*, dan *Rapidminer*.

1.4. TUJUAN DAN MANFAAT PENELITIAN

1.4.1. Tujuan Penelitian

Adapun tujuan penelitian yang dilakukan oleh penulis, yaitu:

1. Menganalisis hasil perhitungan *Clustering* data penjualan dengan algoritma *K-Means Clustering*.
2. Menerapkan algoritma *K-Means Clustering* pada transaksi jual beli di Agen Buah Yudi.

1.4.2. Manfaat Penelitian

Adapun tujuan penelitian yang dilakukan oleh penulis, yaitu:

1. Dengan perhitungan *Cluster* yang akurat maka pengambilan keputusan dalam menentukan prioritas pembeli di Agen Buah Yudi.
2. Penulis dapat menambah ilmu dan wawasan baru mengenai penerapan *Data mining* untuk *Clustering* data penjualan.

1.5. LANDASAN TEORI

1.5.1. *Data mining*

Data mining adalah adalah proses menemukan pola yang menarik dan pengetahuan dari sejumlah besar data atau *data mining* adalah istilah yang mengacu pada penggunaan algoritma dan komputer untuk menemukan novel dan pola menarik dalam data. Menurut ada beberapa jenis fungsionalitas database yang dapat yang dapat dilakukan dalam pengolahan data antara lain adalah [5]:

1. *Class/Concept* Description: *Chara-cterization and Discrimination* (Deskripsi kelas/konsep : karakterisasi dan diskriminasi) : Entri data yang dapat dikaitkan dengan kelas atau konsep. Fungsi ini terdiri dari :
   a. *Data Characterization* (Karakterisasi Data) : adalah ringkasan dari karakteristik umum atau fitur dari kelas target data.
   b. *Data Discrimination* (Diskriminasi Data) : adalah perbandingan fitur umum dari objek data kelas sasaran terhadap fitur umum objek dari satu atau beberapa kelas yang kontras.
2. *Mining Frequent Patterns Associations, and Correlations* (Penggalian pola yang sering muncul : asosiasi dan korelasi) : Meneliti pola yang sering terjadi di data. Fungsi terdiri dari :
   a. *Associations* (Asosiasi) : Pola dimana suatu variabel memiliki confidence (tingkat keyakinan) dengan variabel lain dan support (tingkat pendukung) dimana variabel memiliki pola yang sama.
   b. *Correlations* (Korelasi) : Tingkat hubungan yang dimiliki oleh suatu variabel dengan variabel lain.
3. *Classifications and Regression for Predictive Analysis* (Klasifikasi dan regresi untuk analisis prediksi) :
   a. *Classification* (Klasifikasi) : adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep.
   b. *Regression* (Regresi) : adalah proses untuk mengestimasi hubungan antara variabel.

4. *Cluster analysis* (Analisis klaster) : pengelompokan analisis objek data tanpa konsultasi label kelas.

1.5.2. *K-Means*

Algoritma *K-Means* merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (*centroid*) terdekat dengan data. Tujuan dari *K-Means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik *centroid*.

Tahapan awal yang dilakukan pada proses klasterisasi data dengan menggunakan algoritma *K-Means* adalah pembentukan titik awal *centroid* cj Pada umumnya pembentukan titik awal *centroid* dibangkitkan secara acak. Jumlah *centroid* cj yang dibangkitkan sesuai dengan jumlah klaster yang ditentukan di awal. Setelah k *centroid* terbentuk kemudian dihitung jarak tiap data xi dengan *centroid* ke-j sampai k, dinotasikan dengan d (xi,cj). Terdapat beberapa ukuran jarak yang digunakan sebagai ukuran kemiripan suatu *instance data*, salah satunya adalah jarak *Euclidee*. Perhitungan jarak *Euclideean* seperti dibawah ini :

$$d(Xi, Cj) = \sqrt{\sum_{i=1}^{N} (Xi - Ci)^2}$$

Jika $d(Xi, Cj)$ semakin kecil, kesamaan antara dua unit pengamatan semakin dekat. Syarat menggunakan jarak *Euclide* adalah jika semua fitur dalam dataset tidak saling berkorelasi. Jika terdapat fitur yang berkorelasi maka menggunakan konsep jarak Mahalanobis.

Kelanjutan dari jarak tersebut dicari yang terdekat sehingga data akan mengelompok berdasarkan *centroid* yang paling dekat. Tahap berikutnya adalah update titik *centroid* dengan menghitung rata-rata jarak seluruh data terhadap *centroid*. Selanjutnya akan kembali lagi ke proses awal. Iterasi ini

akan diulangi terus sampai didapatkan *centroid* yang konstan artinya titik *centroid* sudah tidak berubah lagi. Atau iterasi dihentikan berdasarkan jumlah iterasi maksimal yang ditentukan[6].

1.5.3. *Clustering*

Clustering adalah pengelompokan menggunakan teknik *unsupervised* learning dimana tidak diperlukan fase learning serta tidak menggunakan pelabelan pada setiap kelompok. Metode *Clustering* mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama. Tujuan dari *Clustering* ini adalah untuk meminimalisasi fungsi tujuan yang ditetapkan dalam proses *Clustering*, yang umumnya berusaha meminimalisasi variasi dalam suatu *cluster* dan memaksimalisasi variasi antar *cluster*.

Clustering adalah proses pembentukan kelompok data (*cluster*) dari himpunan data yang tidak diketahui kelompok-kelompok atau kelaskelasnya dan proses menentukan data-data termasuk ke dalam *cluster* yang mana. *Clustering* merupakan proses untuk mengetahui kelas-kelas taksonomi atau batryologi, atau analisis topologi dari data-data yang ada. Dilihat dari kacamata *data mining*, *Clustering* bukanlah proses klarifikasi. Karena dalam proses klarifikasi, data dikelompokkan ke dalam kelas-kelas yang telah diketahui sebelumnya[7]

1.5.4. *WEKA*

WEKA adalah sebuah paket tools machine learning praktis. *WEKA* merupakan singkatan dari *Waikato Environment for Knowledge Analysis*, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. *WEKA* mampu menyelesaikan masalah-masalah *data mining* di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan machine learning. Perangkat lunak ini ditulis dalam hirarki class Java dengan metode berorientasi objek dan dapat berjalan hampir di semua platform[8]

WEKA (*Waikato Environment for Knowledge Analysis*) merupakan aplikasi *data mining* yang bersifat open source berbasis Java. *WEKA*

pertama kali dikembangkan oleh Universitas Waikato Selandia Baru sebelum menjadi bagian di Pentanho. *Weka* terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data Karena *Weka* ditulis dalam bahasa pemrograman *Java*, maka *Weka* juga didukung oleh GUI yang sangat baik dan user friendly, dapat mengolah berbagai file data seperti *.csv dan .*arff serta memiliki fitur utama seperti data *preprocessing tools, learning algorithms* dan berbagai metode evaluasi[9].

1.5.5. *Rapidminer*

Rapidminer adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. *Platfrom* dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran. *Rapidminer* memiliki sekitar 100 solusi pembelajaran untuk pengelompokan, klasifikasi dan analisis regresi. *Rapidminer* juga mendukung sekitar 22 format file, seperti .xls, .csv, dan sebagainya[10]

1.6. METODOLOGI PENELITIAN

a. Alat dan Bahan Penelitian

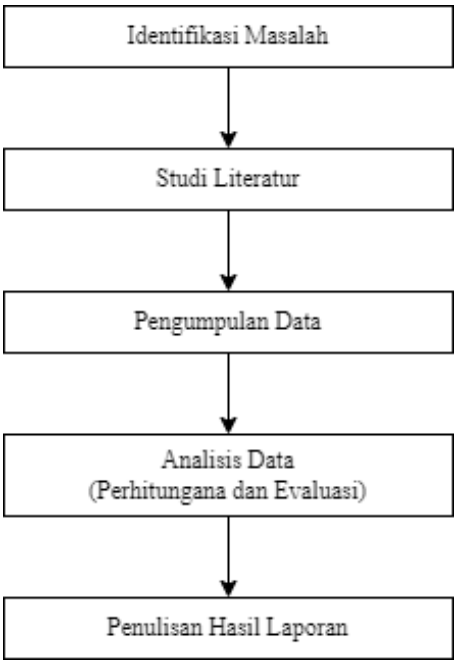
Dalam menganalisis dan menerapkan algoritama apriori untuk penempatan obat pada Apotek Persijam Jambi yang digunakan oleh penulis terdiri dari

1. Perangkat Keras (*Hardware)* yang terdiri dari :
   a) *Notebook* Lenovo Ideapad 320-14AST dengan spefikasi : *Processor* AMD A4-9120
   b) 8GB DDR4 *Memory*
   c) 500GB SSD.

2. Perangkat lunak (*software*) yang terdiri dari :
    a) Sistem Operasi *Windows 11*
    b) *Microsoft Office 2021*
    c) *Notepad*
    d) *Tools WEKA* 3.8.5
    e) *Tools RapidMiner*

b. Metode Penelitian

Agar penelitian berjalan dengan lancar, penulis menggunakan kerangka penelitian yang menjelaskan langkah-langkah yang diperlukan untuk mengatasi masalah penelitian ini, agar penelitian ini terstruktur dan dapat selesai tepat waktu. Kerangka kerja penelitian yang digunakan dapat dilihat pada gambar 1.1 berikut :



Gambar 1. 1 Kerangka Kerja Penelitian

Berdasarkan dari langkah-langkah penelitian di atas, maka dapat dijelaskan tahapan-tahapan dalam penelitian di atas sebagai berikut :

1. Identifikasi Masalah

   Identifikasi masalah merupakan langkah awal yang dilakukan. Hal ini dilakukan untuk menentukan penelitan apa yang akan kita angkat untuk diteliti.

2. Studi Literatur

   Pada tahapan ini, penulis melakukan kajian pustaka yang bermaksud untuk mencari, mempelajari dan mengamati refrensi-refrensi, jurnal dan lain sebagainya yang serupa atau relevan dalam penelitian yang akan penulis teliti. Studi literatur ini bertujuan untuk mendapatkan landasan teoritis mengenai permasalahan yang akan diteliti, yang bertujuan untuk dapat memahami permasalahan yang akan diteliti.

3. Pengumpulan data

   Merupakan tahapan dalam melakukan pengumpulan data dan informasi yang akan penulis teliti dalam penerapan metode *K-Means Clustering* pada Agen Buah Yudi, agar dapat tercapainya suatu tujuan dalam penelitian ini .

4. Analisis Data

   Merupakan tahapan untuk memahami dan mempelajari data-data yang ada agar mempermudahkan penulis untuk melakukan tahap selanjutnya. Dalam tahap ini penulis menganalisi permasalahan yang ada diolah serta dijelaskan lebih detail dalam pembahsan penelitian.

5. Penulisan Laporan

   Pada tahap ini dilakukan penulisan laporan dari penelitian yang bertujuan dapat mengetahui sasaran pembeli dengan metode *K-Means Clustering.*

1.7. JADWAL PENELITIAN

Adapun jadwal penelitian yang dilakukan untuk menyelesaikan penelitian ini adalah :

Tabel 1. 1 Jadwal Penelitian

No.	Kegiatan	Bulan															
		September				Oktober				November				Desember			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Identifikasi Masalah																
2.	Studi Literatur																
3.	Pengumpulan Data																
4.	Analisis Data (perhitungan dan evaluasi)																
5.	Penulisan Laporan																

DAFTAR PUSTAKA

[1] W. Dhuhita, "*Clustering* Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.

[2] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode *K-Means* Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, p. 13, 2021, doi: 10.30872/jim.v16i1.5182.

[3] I. Sumadikarta and E. Abeiza, "PENERAPAN ALGORITMA *K-MEANS* PADA *DATA MINING* UNTUK MEMILIH PRODUK DAN PELANGGAN POTENSIAL (Studi Kasus : PT Mega Arvia Utama)," *J. Satya Inform.*, no. 1, pp. 1–12, 2014.

[4] G. N. W. Paramartha, D. E. Ratnawati, and A. W. Widodo, "Analisis Perbandingan Metode *K-Means* Dengan Improved Semi-Supervised Analisis Perbandingan Metode *K-Means* Dengan Improved Semi-Supervised *K-Means* Pada Data Indeks Pembangunan Manusia (IPM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. Vol. 1, no. 9, pp. 813–824, 2017, [Online]. Available: http://j-ptiik.ub.ac.id

[5] H. T. SIHOTANG, "Sistem Informasi Pengagendaan Surat Berbasis Web Pada Pengadilan Tinggi Medan," vol. 3, no. 1, pp. 6–9, 2019, doi: 10.31227/osf.io/bhj5q.

[6] A. Asroni and R. Adrian, "Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *Weka* Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016, doi: 10.18196/st.v18i1.708.

[7] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma *K-Means Clustering* untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi:

10.26418/jp.v5i1.29611.

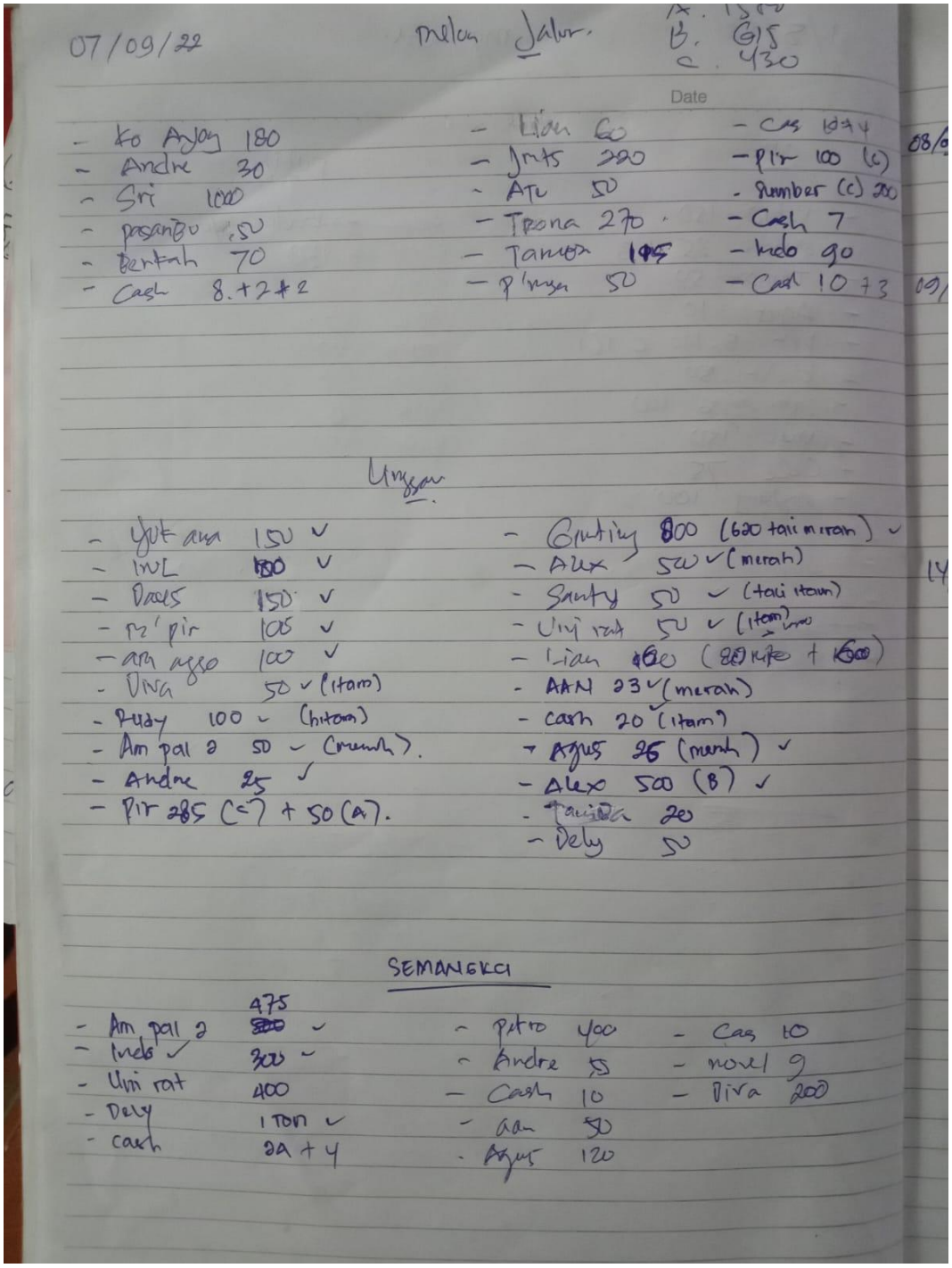
[8] S. Pujiono, A. Amborowati, and M. Suyanto, “Analisis Kepuasan Publik Menggunakan *Weka* Dalam Mewujudkan Good Governance Di Kota Yogyakarta,” *Data Manaj. dan Teknol. Inf.*, vol. 14, no. 2, p. 45, 2013.

[9] T. M. Tamtelahitu, “Komparasi Algoritma *Clustering* dengan Dataset Penyebaran Covid-19 di Indonesia Periode Maret-Mei 2020,” *J. Teknol. Technoscientia*, vol. 13, no. 1, pp. 27–34, 2020.

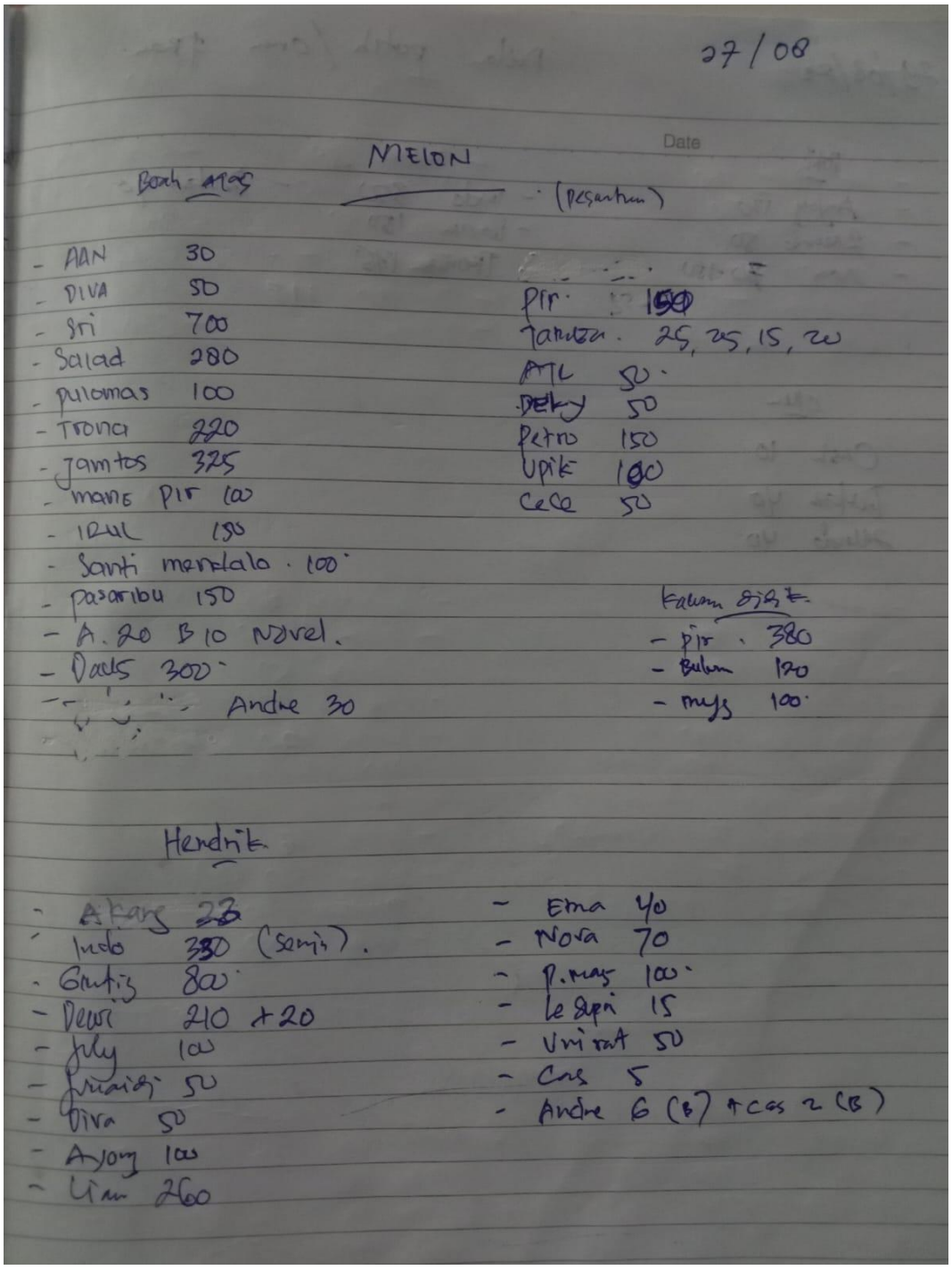
[10] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, “Penerapan Aplikasi *Rapidminer* Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.

LAMPIRAN

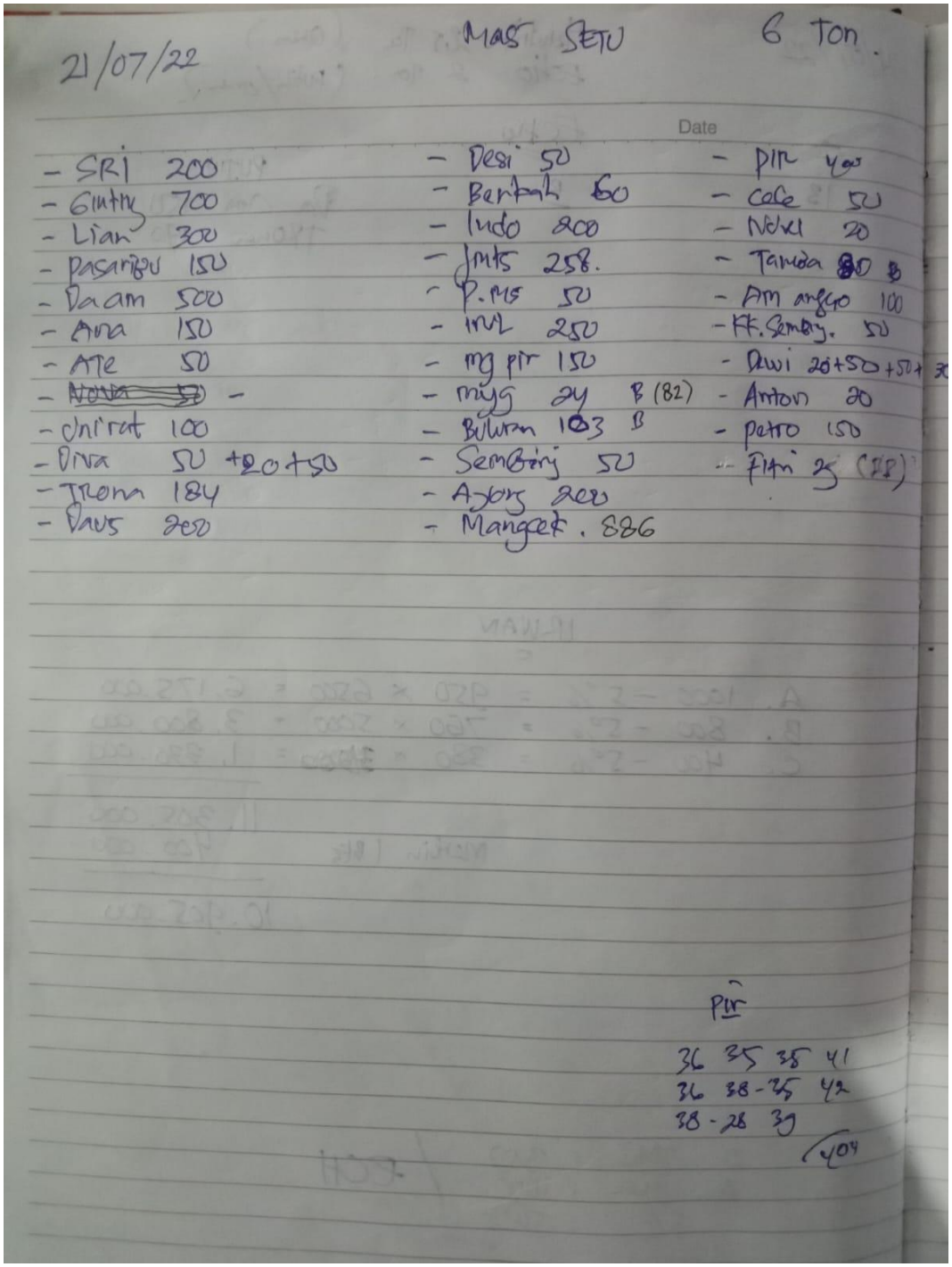
Lampiran data di Agen Buah Yudi masih manual, berikut ini adalah lampiran data tersebut :



Gambar 1. 2 Data Agen Buah Yudi 07/09/22



Gambar 1. 3 Data Agen Buah Yudi 27/08/22



Gambar 1. 4 Data Agen Buah Yudi 21/07/22

Link data Agen Buah Yudi :

https://drive.google.com/drive/folders/1CoRxhHKXVgry2RsCJyLh1_tZiriuCg2f