

## PENERAPAN FUZZY C – MEANS MEMPREDIKSI KEUNTUNGAN DAN PAJAK CUKAI PENJUALAN ROKOK BERDASARKAN KELOMPOK PENGGUNA ROKOK DAN TEMBAKAU

Tia Rizkika<sup>1</sup>, Dicky Apdillah<sup>2</sup>

Universitas Asahan, Asahan

e-mail: <sup>1</sup>tiarizkika1312@gmail.com, <sup>2</sup>dickyapdi1404@gmail.com

**Abstract:** *The cigarette industry is a sector that contributes significantly to state revenue through taxes and excise. This study aims to apply the Fuzzy C-Means method to cluster cigarette consumption and sales data to predict potential sales profits and projected excise revenues. The data used is secondary data obtained from the Central Statistics Agency (BPS) and the Ministry of Finance. The research process includes data collection, pre-processing, analysis using the Fuzzy C-Means algorithm, and implementation of a web-based system using PHP and MySQL. The results show that the Fuzzy C-Means method is capable of grouping cigarette consumption data into several clusters, such as "very profitable," "quite profitable," and "less profitable." This model can assist industry and government players in formulating pricing strategies, excise policies, and production planning based on more accurate data.*

**Keywords:** *Fuzzy C-Means, Clustering, Profit Prediction, Data Mining, PHP, MySQL.*

**Abstrak:** Industri rokok merupakan salah satu sektor yang memberikan kontribusi besar terhadap penerimaan negara melalui pajak dan cukai. Penelitian ini bertujuan untuk menerapkan metode Fuzzy C-Means dalam melakukan pengelompokan (clustering) terhadap data konsumsi dan penjualan rokok guna memprediksi potensi keuntungan penjualan serta proyeksi penerimaan cukai. Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) dan Kementerian Keuangan. Proses penelitian meliputi tahapan pengumpulan data, pra-pemrosesan, analisis dengan algoritma Fuzzy C-Means, serta implementasi sistem berbasis web menggunakan PHP dan MySQL. Hasil penelitian menunjukkan bahwa metode Fuzzy C-Means mampu mengelompokkan data konsumsi rokok menjadi beberapa klaster, seperti kelompok "sangat menguntungkan", "cukup menguntungkan", dan "kurang menguntungkan". Model ini dapat membantu pelaku industri dan pemerintah dalam merumuskan strategi penetapan harga, kebijakan cukai, serta perencanaan produksi berbasis data yang lebih akurat.

**Kata Kunci:** *Fuzzy C-Means, Clustering, Prediksi Keuntungan, Data Mining, PHP, MySQL.*

### PENDAHULUAN

Salah satu industri yang memberikan kontribusi cukup besar terhadap penerimaan negara, khususnya melalui pemungutan cukai, adalah industri rokok. Cukai rokok terus meningkat dari tahun ke tahun, hingga akhirnya menjadi sumber utama penerimaan negara di sektor pajak tidak langsung. Dimulai dari latar belakang pertimbangan, modifikasi

sistem, dan peta jalan penyederhanaan tarif cukai hasil tembakau, proses penetapan tarif cukai cukup rumit dan melibatkan banyak pihak. Cukai hasil tembakau dan pajak pertambahan nilai hasil tembakau merupakan sumber utama penerimaan negara, dan rokok merupakan penyumbang terbesar penerimaan tersebut (Yusuf dkk., 2023).

Data mining adalah pengumpulan dan pemanfaatan data historis untuk

tujuan menemukan keteraturan, pola, atau hubungan dalam basis data yang besar (Edo, 2021). data mining (DM) adalah gabungan sejumlah disiplin ilmu computer ), yang mendefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistic* dan database systems Data mining atau penambahan kata adalah teknik yang relatif cepat dan mudah untuk menemukan pola dan/atau relasi antar data, secara otomatis (Suliman, 2021).

Industri rokok di Indonesia merupakan salah satu sektor yang memberikan kontribusi besar terhadap penerimaan negara melalui pajak dan cukai. Pajak dan cukai rokok memiliki peran penting dalam pengendalian konsumsi rokok serta sebagai sumber pendapatan negara. Besaran tarif pajak dan cukai tersebut telah ditetapkan secara regulatif oleh pemerintah dan mengalami perubahan sesuai dengan kebijakan fiskal yang berlaku.

Di sisi lain, tingkat penjualan rokok dipengaruhi oleh berbagai faktor seperti harga, kelompok pengguna rokok, jenis tembakau, dan tren konsumsi masyarakat. Kondisi ini menyebabkan perusahaan perlu melakukan analisis dan prediksi keuntungan secara cermat untuk dapat menyesuaikan strategi produksidan penjualan.

Algoritma Fuzzy C-Means Penggunaan Algoritma Fuzzy C-Means bertujuan untuk melihat variasi hasil kluster yang dihasilkan dari algoritma yang dilakukan implementasi pada aplikasi yang dibuat berdasarkan pada banyaknya jumlah data yang tersedia. (Damayanti, 2023) Model pengelompokan fuzzy digunakan oleh FCM untuk memastikan bahwa data dapat menjadi anggota dari semua kelas atau klaster yang dihasilkan dengan berbagai tingkat keanggotaan mulai dari 0 hingga Tingkat keanggotaan dalam kelas atau klaster menentukan tingkat data yang ada di kelas atau klaster tersebut. (Nurmin dkk, 2022).

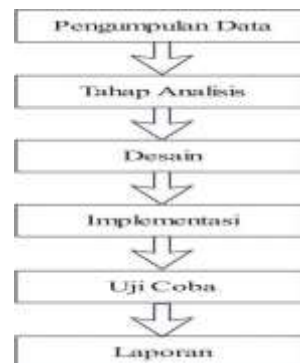
Metode *Fuzzy C- Means* dapat

digunakan untuk melakukan pengelompokan (*clustering*) terhadap data penjualan dan perilaku pengguna rokok. Dengan penerapan metode ini, perusahaan dapat mengetahui pola penjualan berdasarkan kelompok pengguna, sehingga dapat dilakukan prediksi potensi keuntungan pada masing-masing kelompok.

Dalam konteks ini, pajak dan cukai berfungsi sebagai informasi pendukung dalam analisis keuntungan, bukan sebagai variabel yang dihitung secara otomatis oleh sistem, karena nilainya telah ditetapkan oleh pemerintah. Dengan demikian, penelitian ini berfokus pada penerapan metode *Fuzzy C-Means* untuk memprediksi keuntungan penjualan rokok berdasarkan kelompok pengguna, dengan mempertimbangkan pajak dan cukai sebagai faktor eksternal yang mempengaruhi besarnya keuntungan.

## METODE

Dalam melakukan suatu penelitian, peneliti harus melakukan perencanaan untuk melakukan suatu penelitian dapat dilakukan secara tepat dan sistematis. Desain penelitian adalah suatu rencana dari analisis, dan interpretasi yang sama, terperinci, dan spesifik. Rencana penelitian merupakan penjelasan apa yang dilakukan dan masalah apa yang diselesaikan yang ada pada saat penelitian dilakukan. Dengan adanya perencanaan dalam penelitian memungkinkan peneliti dapat mengambil tindakan dan keputusan yang tepat ketika berhadapan dengan masalah penelitian.



Gambar 1 Gambar Kerangka Kerja

Metode-metode kuantitatif tersebut dimanfaatkan dalam metode penelitian ini. Metode kuantitatif ini memanfaatkan data berupa data numerik atau angka-angka yang sifatnya numerik. Untuk tujuan menemukan solusi dari suatu masalah, rumus-rumus matematika dimanfaatkan dalam penelitian yang menggunakan metodologi kuantitatif. Dalam proses pelaksanaan penelitian ini, penulis menggunakan metode pengumpulan data yang dikenal dengan clustering, yang berbasis fuzzy C-Means. Fuzzy C-Means terlebih dahulu akan menentukan pusat cluster, dan pusat cluster ini kemudian akan mengelompokkan setiap titik ke dalam suatu kelompok. Untuk memperoleh pusat cluster, yang kemudian akan bergerak ke posisi yang sesuai karena kondisi di pusat cluster dan setiap titik data berdasarkan derajat keanggotaan, maka penentuan pusat cluster dan nilai keanggotaan akan dilakukan beberapa kali. Hal ini akan memunculkan pusat cluster.

## HASIL DAN PEMBAHASAN

### Hasil Penelitian

Hasil penelitian ini diperoleh melalui penerapan metode *Fuzzy C-Means* untuk mengelompokkan data konsumsi rokok dan tembakau yang bersumber dari Badan Pusat Statistik (BPS) dan laporan resmi Kementerian Keuangan. Fokus penelitian adalah pada rokok putih (Marlboro) karena produk ini menjadi salah satu komoditas penting dalam perhitungan keuntungan penjualan dan pajak cukai. Dengan pengelompokan data, pola konsumsi masyarakat dapat diketahui secara lebih rinci, sehingga hasil analisis dapat memberikan gambaran yang lebih akurat terkait potensi keuntungan dan proyeksi penerimaan negara.

Proses penelitian dilakukan secara bertahap mulai dari pengumpulan data, analisis pola konsumsi, hingga pengujian metode *Fuzzy C-Means*. Hasil yang diperoleh kemudian diimplementasikan

ke dalam sistem berbasis web dengan dukungan PHP dan MySQL sebagai teknologi utama. Dengan cara ini, data yang semula bersifat mentah dapat diproses secara otomatis menjadi informasi yang berguna untuk mendukung pengambilan keputusan, baik bagi pemerintah maupun pelaku industri rokok.

### Analisis Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang telah dipublikasikan secara resmi, sehingga validitasnya dapat dipertanggungjawabkan.

Variabel yang dianalisis terdiri atas enam kategori utama, yaitu rokok kretek filter, rokok kretek tanpa filter, rokok putih, tembakau, serta rokok dan tembakau lainnya. Variabel ini dipilih untuk menggambarkan keberagaman produk rokok di pasaran serta variasi konsumsi masyarakat. Penggunaan variabel yang beragam memungkinkan proses *clustering* menghasilkan pengelompokan data yang lebih representatif.

Sebelum digunakan dalam proses pengolahan, data melewati tahap pra-pemrosesan yang meliputi seleksi, pembersihan, dan transformasi. Seleksi dilakukan untuk memastikan data sesuai dengan kebutuhan penelitian, pembersihan dilakukan untuk menghilangkan data yang tidak konsisten, sedangkan transformasi dilakukan agar data memiliki format yang seragam. Tahapan ini penting agar hasil pengelompokan menggunakan *Fuzzy C-Means* dapat diandalkan serta meminimalisasi kesalahan perhitungan yang mungkin muncul akibat data yang tidak siap digunakan.

### Proses Perhitungan *Fuzzy C-Means*

Proses klasifikasinya melibatkan perhitungan probabilitas awal (prior) dan probabilitas kondisi (likelihood) untuk menentukan probabilitas akhir (posterior) dari masing-masing kelas. Kelas dengan nilai posterior tertinggi akan menjadi hasil prediksi. Karena itu, setiap periode

dan nilai kategorinya harus didefinisikan secara jelas melalui tabel kriteria yang digunakan sebagai dasar dalam proses

klasifikasi. Adapun Langkah-Langkahnya adalah sebagai berikut:

#### Data awal

**Tabel 1 Data Awal**

Nama	Rokok Kretek Filter	Rokok Kretek Tanpa Filter
Simeulue	10,138	3,929
Aceh Singkil	13,897	3,944
Aceh Selatan	13,161	6,398
Aceh Tenggara	14,658	6,648
Aceh Timur	15,686	3,026
Aceh Tengah	14,471	8,128
Aceh Barat	12,054	3,794
Aceh Besar	7,59	4,062

Pada tabel diatas terdapat  $n=8$  sampel (daerah) dan  $p - 2$  atribut maka hal yang pertama adalag menentukan parameter dengan jumlah cluster 3 (tiga). Max Iterasi 100, Pembobotan 2 dan epsilon 0,00001.

Pada iterasi pertama, dilakukan perhitungan nilai keanggotaan awal ( $U$ ) terhadap masing-masing cluster ( $C1, C2, C3$ ). Nilai keanggotaan tersebut kemudian dikuadratkan untuk memperoleh nilai Miu Kuadrat, yang selanjutnya digunakan dalam perhitungan pusat cluster (centroid) sementara. Adapun perhitungannya adalah sebagai berikut:

$$\text{Rumus Miu Kuadrat} = (\mu_{ij})^2 \\ = (0,6)^2 = 0,36$$

Lakukan perhitungan berulang untuk seluruh data pada cluster 1, cluster 2 dan cluster 3 kemudian menggunakan hasil total untuk menghitung centroid  $C1$  pada ietrasi , adapun hasil perhitungan sebelumnya adalah sebagai berikut:

Menghitung Pusat Cluster Awal (Centroid)

$$\sum^n (U_{ij})^m . X_i$$

Langkah selanjutnya adalah menghitung pusat cluster X-V pada metode Fuzzy C-Means (FCM), adalah:

$$D_{1,1} = \sqrt{(10,138 - 12,28153)^2 \times (3,929 - 4,765071)^2} = 5,293735784$$

$$D_{2,1} = \sqrt{(10,138 - 13,69537048)^2 \times (3,929 - 5,061371944)^2} = 13,93715096$$

$$D_{3,1} = \sqrt{(10,138 - 12,52837125)^2 \times (3,929 - 5,139422507)^2} = 7,178997355$$

Menghitung Tabel Jarak Dengan Rumus :

$$L_{1.1} = (x_1 \times v_{j1}) = 0,36 \times 5,293735784 = 1,905744882$$

$$L_{2.1} = (x_1 \times v_{j1}) = 0,017689 \times 13,93715096 = 0,246534263$$

$$L_{3.1} = (x_1 \times v_{j1}) = 0,071289 \times 7,178997355 = 0,511783542$$

Selanjutnya, dilakukan perhitungan serupa untuk data dan cluster lainnya hingga seluruh nilai derajat keanggotaan ( $\mu$ ) diperbarui. Proses ini berlanjut ke iterasi berikutnya dengan menggunakan nilai  $\mu$  baru sebagai acuan dalam menghitung pusat cluster (centroid) yang diperbarui. Iterasi dilakukan secara berulang hingga nilai perubahan keanggotaan antar iterasi menjadi sangat kecil atau telah memenuhi batas konvergensi yang ditentukan. Dengan demikian, proses Fuzzy C-Means berhenti ketika sistem telah mencapai kondisi stabil, dan setiap data telah memiliki nilai keanggotaan terbesar terhadap salah satu cluster yang menjadi dasar penentuan hasil akhir pengelompokan. Adapun hasil akhir dari perhitungana adalah:

**Tabel 4 Hasil Akhir Perhitungan**

No	Kode	Nama Kabupaten/Kota	Cluster	Status
----	------	---------------------	---------	--------

1	A001	Simeulue	C3	Kurang Menguntungkan
2	A002	Aceh Singkil	C3	Kurang Menguntungkan
3	A003	Aceh Selatan	C2	Cukup Menguntungkan
4	A004	Aceh Tenggara	C2	Cukup Menguntungkan
5	A005	Aceh Timur	C2	Cukup Menguntungkan
6	A006	Aceh Tengah	C2	Cukup Menguntungkan
7	A007	Aceh Barat	C3	Kurang Menguntungkan
8	A008	Aceh Besar	C1	Sangat Menguntungkan
9	A009	Pidie	C1	Sangat Menguntungkan
10	A010	Bireuen	C1	Sangat Menguntungkan
11	A011	Aceh Utara	C1	Sangat Menguntungkan
12	A012	Aceh Barat Daya	C2	Cukup Menguntungkan
13	A013	Gayo Lues	C2	Cukup Menguntungkan
14	A014	Aceh Tamiang	C3	Kurang Menguntungkan
15	A015	Nagan Raya	C3	Kurang Menguntungkan
16	A016	Aceh Jaya	C2	Cukup Menguntungkan
17	A017	Bener Meriah	C2	Cukup Menguntungkan
18	A018	Pidie Jaya	C3	Kurang Menguntungkan
19	A019	Kota Banda Aceh	C3	Kurang Menguntungkan
20	A020	Kota Sabang	C3	Kurang Menguntungkan
21	A021	Kota Langsa	C1	Sangat Menguntungkan
22	A022	Kota Lhokseumawe	C1	Sangat Menguntungkan
23	A023	Kota Subulussalam	C3	Kurang Menguntungkan
24	A024	Nias	C1	Sangat Menguntungkan

## Pembahasan

Pembahasan pada bab ini menguraikan keterkaitan hasil penelitian dengan rumusan masalah serta landasan teori yang telah dibahas pada bab sebelumnya. Hasil pengelompokan menggunakan metode *Fuzzy C-Means* menunjukkan bahwa data konsumsi rokok dan tembakau dapat dikelompokkan ke dalam beberapa cluster dengan tingkat keanggotaan yang berbeda. Hal ini sesuai dengan karakteristik algoritma FCM yang mampu memberikan fleksibilitas dalam klasifikasi data sehingga pola konsumsi dapat dipahami lebih detail dibandingkan metode pengelompokan konvensional. Selain itu, hasil yang diperoleh mendukung teori bahwa pemodelan berbasis data mining, khususnya metode clustering, mampu memberikan kontribusi signifikan dalam pengambilan keputusan. Dengan adanya sistem ini, informasi terkait prediksi keuntungan dan pajak cukai tidak hanya dapat dipantau, tetapi juga dapat digunakan sebagai dasar penyusunan kebijakan cukai maupun

strategi pemasaran industri rokok

## Halaman Utama

Halaman utama adalah tampilan pertama yang muncul saat pengguna membuka sistem. Halaman ini dirancang untuk memberikan akses langsung ke fitur-fitur utama sistem dengan tampilan yang sederhana, jelas, dan mudah dinavigasi.



Gambar 2 Halaman Utama

## Halaman Perhitungan *Fuzzy C-Means*

Halaman Perhitungan *Fuzzy C-Means* adalah tampilan yang muncul saat user atau pengguna memilih menu Perhitungan *Fuzzy C-Means* pada halaman utama sistem tanpa perlu melakukan *login* kedalam sistem.



**Gambar 3 Halaman Perhitungan Fuzzy C-Means**

### Halaman Utama Admin

Halaman utama Admin adalah antarmuka yang ditampilkan setelah Admin berhasil melakukan login dan mengakses sistem. Halaman ini memberikan akses penuh kepada Admin untuk mengelola dan memantau berbagai fitur dalam sistem. Di halaman utama ini, Admin dapat mengakses menu-menu penting seperti data periode, data kriteria, data kabupaten / kota, Perhitungan FUZZY C-MEANS, serta opsi ganti password dan logout.



**Gambar 4 Halaman Utama Admin**

### SIMPULAN

Berdasarkan penelitian yang dilakukan penulis dalam Penerapan *Fuzzy C-Means* Memprediksi Keuntungan Dan Pajak Cukai Penjualan Rokok Berdasarkan Kelompok Pengguna Rokok Dan Tembakau, dapat disimpulkan sebagai berikut.

1. Perhitungan keuntungan pajak cukai penjualan rokok berdasarkan kelompok pengguna rokok dan tembakau dapat dilakukan dengan metode Clustering Fuzzy C-Means. Melalui algoritma ini, data konsumsi rokok diolah menjadi beberapa kelompok dengan tingkat

keanggotaan tertentu, sehingga dapat diketahui pola konsumsi masing-masing cluster berdasarkan kriteria.

2. Pengelompokan jumlah konsumsi rokok berdasarkan kelompok pengguna rokok dan tembakau, khususnya rokok putih, berhasil dilakukan dengan metode Fuzzy C-Means. Hasil clustering menunjukkan adanya perbedaan karakteristik konsumsi antar kelompok, sehingga memudahkan dalam memahami perilaku pasar. Dengan demikian, metode ini mampu memberikan gambaran yang lebih jelas mengenai distribusi konsumsi rokok putih di masyarakat, yang selanjutnya dapat dijadikan dasar dalam menyusun kebijakan cukai maupun strategi pemasaran industri rokok.

### DAFTAR PUSTAKA

- Annisa, K., Serasi Ginting, B., & Syari, M. A. (2022). *Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada Pdam Langkat*.
- Ardi Rahmawan, L. R. P. L. R. (2023). *Sistem Informasi Pemasaran Perumahan Berbasis Web*.
- Azura, A., & Wildian, W. (2018). Rancang Bangun Sistem Absensi Mahasiswa Menggunakan Sensor Rfid Dengan Database Mysql Xampp Dan Interface Visual Basic. *Jurnal Fisika Unand*, 7(2), 186–193. <https://doi.org/10.25077/jfu.7.2.186-193.2018>
- Davvaz, B., Mukhlash, I., & Soleha, S. (2021). Himpunan Fuzzy Dan Rough Sets. *Limits: Journal Of Mathematics And Its Applications*, 18(1), 79. <https://doi.org/10.12962/Limits.V18i1.7705>
- Dedi Irawan, M., & Jend Ahmad Yani Kisaran Sumatera Utara, J. (2018).

- 
- Implementasi Logika Fuzzy Dalam Menentukan Jurusan Bagi Siswa Baru Sekolah Menengah Kejuruan (Smk) Negeri 1  
Air Putih. *Jurnal Teknologi Informasi*, 2  
Ginanjari Perdana P, E. W. B. P. (2023). *Analisis Dampak Kenaikan Tarif Cukai Hasil Tembakau Dan Fenomena Downtrading Rokok Terhadap Penerimaan Negara Dan Produksi Tembakau Domestik 2023*.  
Gohae, F. O., Studi, P., Informatika, T., Teknik, F., Area, U. M., & Area, U.  
M. (2023). *Penerapan Fuzzy C-Means Clustering*.  
Gusman, A. P. (2019). Online Berbasis Costumer Relationship Management (Crm). *Majalah Ilmiah Upi Yptk*, 26(1), 7–13.  
Gusnawan, R., Wijayanto, P., Terapan, F. I., & Telkom, U. (2019). *Aplikasi Kelompok Informasi Masyarakat ( Kim ) Berbasis Web Management Of Community Information Group Application (Kim ) Web- Based 1*