

PENERAPAN METODE CLUSTERING UNTUK MEMETAKAN POTENSI TANAMAN KEDELAI DI JAWA TENGAH DENGAN ALGORITMA FUZZY C- MEANS

Indra Setiawan¹

¹Teknik Informatika Universitas Dian Nuswantoro Semarang

E-mail : 111201005632@mhs.dinus.ac.id

ABSTRAK

Berdasarkan data hasil pertanian kedelai di Dinas Pertanian Provinsi Jawa Tengah, menampilkan beberapa daerah dengan hasil panen kedelai yang bervariasi jumlahnya. Untuk itu diperlukan pengelompokan daerah potensial penghasil kedelai untuk mengetahui daerah mana saja yang menghasilkan kedelai dengan jumlah yang banyak ataupun sedikit. Pembagian hasil panen biasanya dilakukan berdasarkan nama daerah kota / kabupaten penghasil kedelai. Oleh karena itu, dibutuhkan metode untuk memudahkan dalam pengelompokan daerah penghasil kedelai. Dengan pendekatan pengklasteran fuzzy, pembagian kelompok daerah dapat dilakukan berdasarkan luas panen (Ha), hasil per hektar (Kw), produksi dan tahun panen.

Pada penelitian ini dilakukan proses pengklasteran daerah potensial penghasil kedelai menggunakan algoritma Fuzzy C-Means. Dengan menggunakan Fuzzy C-Means bertujuan dalam memudahkan pengelompokan suatu daerah dengan hasil panen kedelai terbesar, sedang dan rendah serta mengukur tingkat akurasi. Hasilnya adalah sebuah gambaran yang menunjukkan pengelompokan daerah berdasarkan hasil pertanian kedelai.

Kata kunci: Pengelompokan daerah penghasil kedelai, Data mining, Cluster, Algoritma Fuzzy C-Means

I. PENDAHULUAN

1.1 Latar Belakang

Kedelai (*Glycine max* (L.) Merrill) merupakan bahan pangan yang mengandung protein nabati yang banyak dikonsumsi oleh penduduk Indonesia. Khususnya di wilayah Jawa Tengah, kedelai biasa diolah menjadi bahan pangan antara lain seperti: tempe, tahu, kecap, susu kedelai, tauco, bungkil kedelai dan sebagainya.

Seiring dengan pertumbuhan penduduk dan berkembangnya industri pengolahan bahan pangan yang menggunakan bahan baku kedelai, maka kebutuhan konsumsi kedelai di dalam negeri terus meningkat. Karena bagi Indonesia kacang kedelai mempunyai peran besar sebagai sumber bahan baku utama bagi industri seperti tahu, tempe dan pakan ternak. Data dari Badan Pusat Statistik (BPS) pada

tabel 1.1 menunjukkan bahwa kebutuhan kacang kedelai dalam negeri cenderung meningkat dan produksi kacang kedelai dalam negeri hanya mampu memenuhi sebagian kecil dari kebutuhan masyarakat Indonesia dalam lima tahun terakhir.

Tabel 1.1 Produksi, impor, ekspor dan kebutuhan dalam negeri kedelai di Indonesia tahun 2006-2010

Tahun	Produksi	Impor	Kebutuhan dlm negeri
2006	747.611	1.132.144	1.878.023
2007	592.534	1.411.589	2.002.251
2008	775.710	1.173.097	1.947.782
2009	974.512	1.314.620	2.288.686
2010	907.031	1.740.505	2.647.151

(ton)

Sumber: BPS(diolah),2011

Untuk itu dalam rangka memenuhi kebutuhan kedelai, dinas pertanian berupaya untuk mengoptimalkan hasil pertanian dengan mengelompokkan daerah yang menghasilkan tanaman kedelai di daerah Jawa Tengah dengan metode *clustering*. Tujuannya adalah untuk mengetahui daerah potensial penghasil kedelai dan dapat mengetahui daerah tersebut cocok untuk tanaman kedelai. Selain itu diperlukan sentralisasi daerah penghasil tanaman kedelai guna memenuhi kebutuhan konsumsi kedelai oleh masyarakat Jawa Tengah agar tidak terjadi kekurangan bahkan mengharuskan import dari luar negeri. Oleh karena itu diperlukan data daerah – daerah penghasil kedelai di Jawa Tengah untuk melakukan pengelompokan dengan algoritma *Fuzzy C-Means*. Dengan data yang sudah dikelompokkan menggunakan algoritma *Fuzzy C-means* diharapkan dapat mempermudah dinas pertanian dalam menghitung hasil pertanian di tiap daerahnya agar mengetahui daerah mana yang menghasilkan kedelai terbanyak, sedang dan sedikit.

Pada metode clustering dibagi menjadi 2 yaitu *hierarchical clustering methods* dan *non hierarchical clustering methods*, metode hirarki digunakan jika jumlah kelompok yang diinginkan ditentukan belum diketahui, sedangkan metode non-hirarki digunakan jika jumlah kelompok yang diinginkan telah ditentukan sebelumnya. *Fuzzy c-means* merupakan contoh non-hirarki yang sering digunakan karena cocok untuk mengolah data berukuran besar dan tipe peubah kontinu [1].

FCM menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau kelompok terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau

cluster ditentukan oleh derajat keanggotaannya [3]. Sebelum diproses menggunakan algoritma *Fuzzy C-Means*, tentukan berapa kelompok yang akan dibentuk atau dibuat terlebih dahulu.

Dari penelitian beberapa pernyataan peneliti sebelumnya, penulis mempunyai ide untuk melakukan pengelompokan daerah pertanian di Jawa Tengah khususnya pada hasil tanaman kedelai dengan menggunakan *Fuzzy C-means Clustering (FCM)*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dikemukakan sebelumnya, maka rumusan masalah dapat dituliskan sebagai berikut:

1. Bagaimana perhitungan akurasi pengelompokan daerah penghasil kedelai menggunakan algoritma *Fuzzy C-Means*.
2. Bagaimana cara mengimplementasikan metode clustering untuk mengelompokkan daerah penghasil kedelai kedalam sebuah perangkat lunak.

1.3 Batasan Masalah

Untuk menghindari penyimpangan dari judul dan tujuan yang sebenarnya serta keterbatasan pengetahuan yang dimiliki penulis, maka penulis membuat ruang lingkup dan batasan masalah yaitu:

1. Pada penelitian ini peneliti akan menggunakan metode *Clustering* dengan algoritma *Fuzzy C-Means* untuk mengelompokkan daerah-daerah penghasil tanaman kedelai di Jawa Tengah.
2. Data yang digunakan untuk pengamatan adalah data hasil panen kedelai wilayah Jawa Tengah dari tahun 2005-2012.
3. Menampilkan pengelompokan dari masing – masing daerah penghasil tanaman kedelai di Jawa Tengah yang terbentuk.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan dari penelitian adalah:

1. Menghitung akurasi hasil perhitungan nilai clustering menggunakan algoritma *Fuzzy C-Means*.
2. Mengimplementasikan model clustering yang dihasilkan untuk mengelompokkan daerah penghasil kedelai kedalam perangkat lunak.

1.5 Manfaat penelitian

Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut:

- a) Menambah pemahaman dan pengalaman dalam menggunakan metode *Clustering* dengan algoritma *Fuzzy C-Means* dalam memetakan atau mengelompokkan suatu data.
- b) Dengan adanya pengelompokan daerah potensi penghasil kedelai di Jawa Tengah, diharapkan hasil pengelompokan daerah ini dapat menjadi acuan daerah mana yang dapat ditanami kedelai.
- c) Membantu dalam pengolahan data dan pengelompokan daerah penghasil pertanian secara cepat dan akurat serta dapat mengetahui daerah-daerah yang dapat menghasilkan tanaman kedelai di Jawa Tengah.

II. TINJUAN PUSTAKA DAN LANDASAN TEORI

2.1 Konsep Data Mining

Konsep dasar data mining adalah menemukan informasi tersembunyi dalam

sebuah basis data dan merupakan bagian dari Knowledge Discovery in Databases (KDD) untuk menemukan informasi dan pola yang berguna dalam data [6]. Data mining mencari informasi baru, berharga dan berguna dalam sekumpulan data dengan melibatkan komputer dan manusia serta bersifat iteratif baik melalui proses yang otomatis ataupun manual.

2.2 Clustering

Pengclusteran merupakan pengelompokkan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Cluster adalah kumpulan record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran.

2.3 Fuzzy Clustering

Fuzzy clustering adalah bagian dari pattern recognition atau pengenalan pola. Fuzzy clustering memainkan peran yang paling penting dalam pencarian struktur dalam data [9]. Fuzzy clustering adalah salah satu teknik untuk menentukan cluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal Euclidian untuk jarak antar vector [8].

2.4 Fuzzy C-Means (FCM)

Merupakan salah satu metode clustering yang merupakan bagian dari metode *Hard K-Means*. Berbeda dengan k-means clustering, dimana suatu objek hanya akan menjadi anggota satu cluster, dalam *Fuzzy C-Means* setiap objek bisa menjadi anggota dari beberapa cluster, sesuai dengan namanya fuzzy yang berarti samar. Batas-batas dalam *K-Means* adalah tegas

(hard) sedangkan dalam *Fuzzy C-Means* adalah soft [10].

Algoritma FCM secara lengkap [3]:

1. Tentukan :
 - a. Matriks X berukuran n x m, dengan n = jumlah data yang akan di cluster; dan m = jumlah variabel (kriteria).
 - b. Jumlah cluster yang akan dibentuk ($C \geq 2$)
 - c. Pangkat (pembobot $w > 1$)
 - d. Maksimum iterasi
 - e. Kriteria penghentian (ϵ = nilai positif yang sangat kecil)
2. Bentuk matriks partisi awal U (derajat keanggotaan dalam cluster); matriks partisi awal biasanya dibuat secara acak.

$$U = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \dots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \dots & \mu_{2n}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{c1}(x_1) & \mu_{c2}(x_2) & \dots & \mu_{cn}(x_n) \end{bmatrix}$$

3. Hitung pusat cluster V untuk setiap cluster

$$V_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^w \cdot x_{kj}}{\sum_{k=1}^n (\mu_{ik})^w}$$

dengan:

V_{ij} : pusat cluster

m : pembobot eksponen

μ_{ik} : fungsi keanggotaan objek ke-k ke gerombol ke-i

X_{kj} : objek ke-k gerombol ke-j

4. Perbaiki derajat keanggotaan setiap data pada setiap cluster (perbaiki matriks partisi)

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(w-1)} \right]^{-1}$$

dengan:

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^m (x_{kj} - v_{ij}) \right]^{1/2}$$

5. Tentukan kriteria penghentian iterasi, yaitu perubahan matriks partisi pada iterasi sekarang dan iterasi sebelumnya

$$\Delta = \|U^t - U^{t-1}\|$$

Apabila $\Delta < \epsilon$ maka iterasi dihentikan.

Ketika kelompok atau setiap objek memungkinkan termasuk ke beberapa cluster, maka μ_{ik} dapat diinterpretasikan sebagai fungsi keanggotaan yaitu $\epsilon [0,1]$. Maka fungsi objektif J yang dirumuskan sebagai fungsi dari U dan V sebagai berikut:

$$J(U, V) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m d^2(x_k v_i)$$

dengan:

U : matriks keanggotaan objek ke masing-masing gerombol

V: matriks centroid / rata-rata masing-masing gerombol

m : pembobot eksponen

μ_{ik} : fungsi keanggotaan objek ke-k ke gerombol ke-i

x_k : objek ke-k

v_i : nilai centroid ke-i

d : ukuran jarak

III. METODE PENELITIAN

3.1 Metode Pengembangan Perangkat Lunak

Pada penelitian ini metode pengembangan perangkat lunak yang digunakan adalah model Rapid Application Development (RAD). Model RAD dipilih karena model proses pengembangan perangkat lunak ini menekankan siklus pengembangan yang memerlukan waktu singkat. Pendekatan RAD meliputi fase-fase berikut ini:

1. Pemodelan Bisnis (*Bussiness Modelling*)

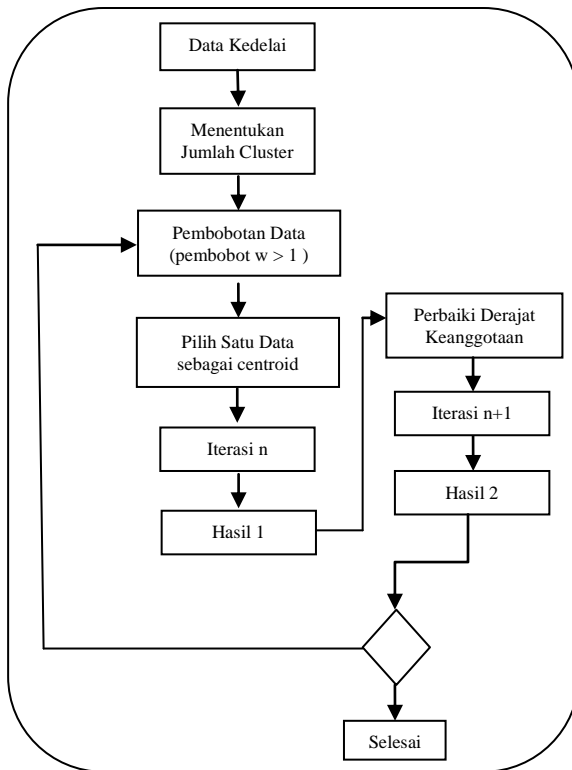
Penelitian ini bertujuan bisnis dalam menentukan aturan clustering yang akan terbentuk dari pengelompokan hasil panen kedelai di wilayah Jawa Tengah sehingga hasil dari pengelompokan dapat mengatur letak dan luas lahan pertanian pada daerah tertentu yg berpotensi menghasilkan kedelai.

2. Pemodelan data (*Data Modelling*)

Pengumpulan data primer yang dibutuhkan dengan mengambil data hasil panen pertanian kedelai yang berupa, nama kabupaten atau kota (namkot), luas panen, hasil panen per hektar, produksi per ton, tahun yang didapat langsung dari pusat data Dinas Pertanian Provinsi Jawa Tengah. Kemudian data dimodelkan agar sistem dapat memproses data tersebut.

3. Pemodelan proses (*Process Modelling*)

Pemodelan proses dalam mengelompokkan menggunakan algoritma *Fuzzy C-Means*



Gambar 2.1 Algoritma Fuzzy C-Means

4. Pembuatan aplikasi (*Application generation*)

Tahap pembuatan program dengan menggunakan bahasa pemrograman PHP. Berikut ini adalah hal yang perlu dilakukan dalam tahap pembuatan program:

- 1) Mendesain alur program

Adalah desain proses bisnis dari program yang akan dibuat.

- 2) Mendesain User Interface (GUI).

Adalah desain User Interface sederhana dari program yang akan dibuat sehingga memudahkan user dalam memahami.

- 3) Memasukan Algoritma Fuzzy C-Means adalah tahap embedding atau penamaan algoritma yang akan memproses analisis data mining.
- 4) Desain Database untuk menyimpan data set hasil pertanian kedelai.
- 5) Melakukan proses perhitungan data secara manual.

5. Pengujian dan pergantian (*Testing and turnover*)

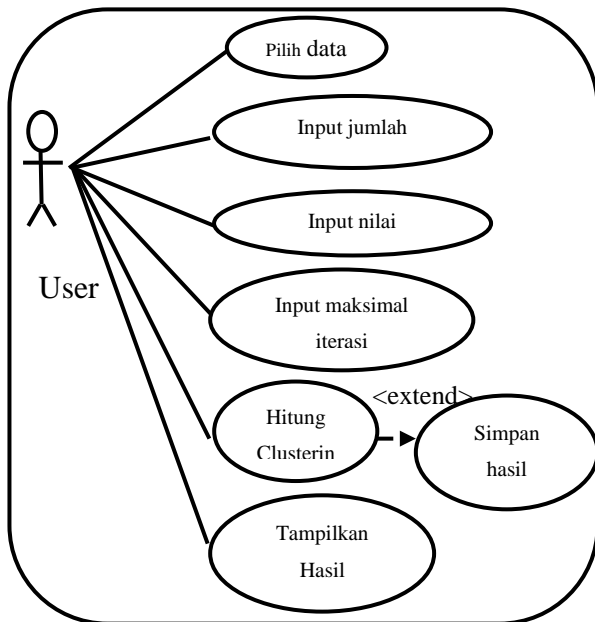
Tahap pengujian dalam penelitian ini meliputi penyempurnaan fungsi program, pencarian dan koreksi kesalahan program pada sintaknya (Syntax Error), pada pengujian ini menggunakan pengujian *Black-box*.

IV. RANCANGAN SISTEM DAN IMPLEMENTASI

Dalam bagian rancangan sistem ini akan dijelaskan mengenai desain yang dibuat berdasarkan langkah-langkah pada metode penelitian dalam mengelompokkan daerah penghasil kedelai dengan algoritma *fuzzy c-means*. Berikut ini adalah langkah yang perlu dikerjakan dalam merancang:

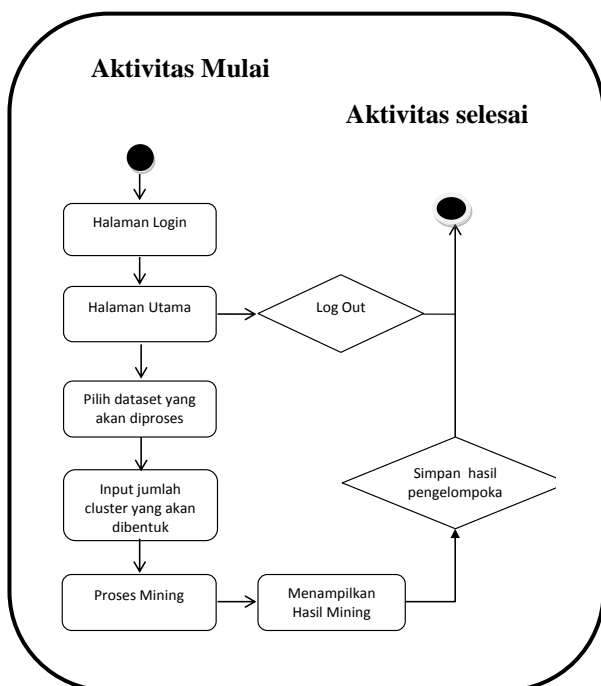
RANCANGAN SISTEM

4.1 Use Case Diagram



Gambar 4.3 Use Case Diagram Rancangan Sistem

4.2 Activity Diagram



Gambar 4.4 Activity Diagram Rancangan Sistem

Activity diagram menunjukkan bagaimana kegiatan pengguna dalam melakukan proses mining. Saat pengguna menjalankan program,

pengguna akan masuk ke halaman login. Selanjutnya pengguna akan masuk ke halaman utama atau logout dari program. Aktifitas pengguna akan berhenti jika memilih logout dari program. Untuk melakukan proses perhitungan pertama pengguna harus melakukan pemilihan dataset berdasarkan tahun yang akan diiproses. Selanjutnya pengguna dapat menentukan inputan yang sesuai mengenai jumlah cluster yang akan dibentuk, pembobotan (pangkat), maupun maksimal iterasinya. Lalu dengan memilih proses maka data akan segera diproses menggunakan algoritma *fuzzy c-means* didalam sistem dan menampilkan hasil mining berupa cluster yang terbentuk.

V. HASIL PENELITIAN DAN PEMBAHASAN

5.1 Hasil Penelitian

Pada hasil penelitian menggunakan data hasil pertanian kedelai di Jawa Tengah yang sudah diolah menjadi dataset pada periode tahun 2005 sampai dengan tahun 2008 dengan menggunakan metode *Fuzzy C-Means* menghasilkan pengelompokan daerah-daerah yang menghasilkan kedelai.

VI. KESIMPULAN DAN PENELITIAN SELANJUTNYA

6.1 Kesimpulan

Setelah melakukan proses percobaan pengelompokan daerah potensial penghasil kedelai dengan menggunakan metode *Fuzzy C-Means* dapat disimpulkan bahwa:

1. Dengan memanfaatkan proses pengolahan data menggunakan konsep data mining, maka proses pengolahan data cara konvensional dapat diatasi.

2. Dengan dibuatnya software pengelompokan data dengan menggunakan algoritma fuzzy c-means yang diakses menggunakan bahasa pemrograman PHP dapat membantu pengguna dalam memproses data hasil pertanian kedelai di Jawa Tengah
3. Hasil percobaan pengelompokan menggunakan software ini diperoleh 3 kelompok atau cluster daerah penghasil tanaman kedelai di Jawa Tengah.

6.2 Penelitian Selanjutnya

Hasil pengelompokan daerah potensial penghasil kedelai menggunakan metode clustering dengan algoritma *Fuzzy C-means* dapat dilakukan penelitian selanjutnya dengan menambah variabel tambahan dan menghubungkan antara daerah penghasil pertanian dengan daerah rawan bencana untuk mengetahui daerah pertanian yang masuk ke daerah rawan bencana dengan metode lain.

DAFTAR PUSTAKA

- [1] Lathifaturramah, "Perbandingan Hasil Pengelompokan Metode K-Means," Institut Pertanian Bogor, Bogor, 2010.
- [2] H. Harianja, "Visualisasi K-Means Clustering pada Data Potensi Pertanian Desa di Bogor menggunakan Mapserver," Institut Pertanian Bogor, Bogor, 2008.
- [3] E. T. Luthfi, "Fuzzy C-Means Untuk Clustering Data (Studi Kasus: Data Performance Mengajar Dosen)," in *Seminar Nasional Teknologi*, Yogyakarta, 2007.
- [4] A. D. Titin Agustin, "Penggerombolan Daerah Tertinggal di Indonesia Dengan Fuzzy C-Means," *Forum Statistika dan Komputasi*, vol. 15 No.1, pp. 22-27, 2010.
- [5] Bahar, "Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C-Means," UDINUS, Semarang, 2011.
- [6] M. H. Dunham, *Data Mining Introductory and Advance Topics*, New Jersey: Prentice Hall, 2003.
- [7] D. T. Larose, *Discovering Knowledge In Data An Introduction to Data Mining*, New Jersey: John Wiley & Sons, Inc, 2005.
- [8] S. P. H. Kusumadewi, *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*, Yogyakarta: Graha Ilmu, 2004.
- [9] "Fuzzy Clustering Algorithms for Unsupervised Change Detection In Remote Sensing Images," *Elsevier*, 2010.
- [10] A. Y, "K-Means Penerapan Permasalahan dan Metode Terkait," *Jurnal Sistem dan Informatika*, vol. III, pp. 47-60, 2007.
- [11] K. Peranginangin, *Aplikasi Web dengan PHP dan MySQL*, Yogyakarta: CV. Andi Offset, 2006, 2006.
- [12] B. Sidik, "MySQL," Bandung, Informatika, 2005, p. 1.
- [13] I. Sommerville, *Software engineering 9th edition*, Boston: Pearson Education Inc, 2011.
- [14] G. B. & H. J. Rosenblatt, *Systems Analysis and Design Edition 9th*, Boston: An imprint of course technology, 2009.
- [15] R. Arindiono, "Perancangan media pembelajaran interaktif matematika untuk siswa kelas 5 SD," *Jurnal Sains dan Seni Pomits*, vol. I, p. 5, 2013.