

**KLASIFIKASI PENYAKIT KANKER PARU-PARU BERDASARKAN
FAKTOR RISIKO MENGGUNAKAN METODE *K-MEANS***

SKRIPSI

Oleh :
HIDA MUHIMMATUL HUSNA
NIM. 18650115



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**KLASIFIKASI PENYAKIT KANKER PARU-PARU BERDASARKAN
FAKTOR RISIKO MENGGUNAKAN METODE *K-MEANS***

SKRIPSI

Oleh:

**HIDA MUHIMMATUL HUSNA
NIM 18650115**

Diajukan Kepada:

**Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

HALAMAN PERSETUJUAN

KLASIFIKASI PENYAKIT KANKER PARU-PARU BERDASARKAN
FAKTOR RISIKO MENGGUNAKAN METODE K-MEANS

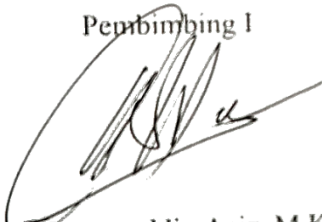
SKRIPSI

Oleh :


HIDA MUHIMMATUL HUSNA
NIM. 18650115

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal : 14 Juni 2023

Pembimbing I


Okta Qomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013

Pembimbing II


Dr. Cahyo Crysdiyan, MCS
NIP. 19740424 200901 1 008


Mengetahui,

Ketua Program Studi Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrud Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI PENYAKIT KANKER PARU-PARU BERDASARKAN
FAKTOR RISIKO MENGGUNAKAN METODE K-MEANS

SKRIPSI

Oleh:

HIDA MUHIMMATUL HUSNA
NIM. 18650115

Telah Dipertahankan di Depan Dewan Penguji
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal: 23 Juni 2023

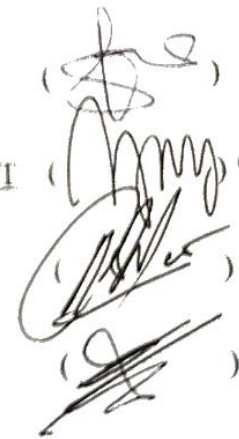
Susunan Dewan Penguji

Ketua Penguji : Dr. M. Amin Hariyadi, M.T
NIP. 19670118 200501 1 001

Anggota Penguji I : Agung Teguh Wibowo Almats, M.T
NIP. 19860301 20180201 1 235

Anggota Penguji II : Okta Qomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013

Anggota Penguji III : Dr. Cahyo Crysdian, MCS
NIP. 19740424 200901 1 008



Mengetahui,

Ketua Program Studi Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Hida Muhiimmatul Husna

NIM : 18650115

Fakultas : Sains dan Teknologi

Jurusan : Teknik Informatika

Judul Skripsi : Klasifikasi Penyakit Kanker Paru-Paru Berdasarkan Faktor Risiko Menggunakan Metode *K-Means*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apaabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 22 Juni 2023
Yang Membuat pernyataan,



Hida Muhiimmatul Husna
18650115

HALAMAN MOTTO

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”

(QS. Al-Baqarah : 285)

HALAMAN PERSEMBAHAN

أَحْمَدُ لِلَّهِ رَبِّ الْعَالَمِينَ

Puji syukur atas kehadiran Allah SWT atas segala rahmat dan karunia-Nya
Shalawat serta salam kepada Nabi Agung Muhammad SAW

Penulis mempersembahkan sebuah karya ini dengan penuh rasa syukur kepada:

Kedua orang tua tercinta serta seluruh anggota keluarga saya selaku orang-orang yang selalu percaya kepada penulis, yang telah memberikan kasih sayang, dukungan baik secara moral maupun material, dan doa selama perjalanan panjang saya dalam menyelesaikan studi ini. Terima kasih, Bapak dan Ibu.

Seluruh anggota keluarga tercinta saya juga selaku orang yang selalu percaya kepada penulis, yang selalu memberikan dukungan, motivasi, dan semangat sehingga penulis dapat menyelesaikan skripsi ini. Terima kasih atas doa dan dukungan tanpa henti.

Dosen pembimbing Bapak Okta Qomaruddin Aziz, M.Kom dan Bapak Dr. Cahyo Crysdiyan, MCS atas bimbingan, arahan, dan waktunya yang berharga dalam membantu saya menyelesaikan skripsi ini. Terimakasih atas kesabaran dan dedikasinya.

Dosen penguji Bapak Dr. M. Amin Hariyadi, M.T dan Bapak Agung Teguh Wibowo Almais, M.T yang telah membimbing dan memberikan masukan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini.

Segenap sivitas akademika Program Studi Teknik Informatika UIN Maulana Malik Ibrahim Malang, terutama seluruh dosen. Terima kasih atas segenap ilmu dan bimbingannya.

Teman-teman Teknik Informatika Angkatan 2018 yang telah memberikan dukungan serta doa, dan menjadi teman diskusi dalam menuntut ilmu. Terima kasih.

Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan kontribusi dan dukungan selama penyusunan skripsi ini.

Penulis ucapkan “Jazaakumullahu khairan katsiiraa, jazaakumullah ahsanal jazaa”. Semoga tali silaturahmi selalu terjaga dan Allah selalu ridho. Aamiin Allahumma Aamiin.

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji syukur kehadiran Allah SWT, yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi ini. Shalawat serta salam semoga tercurah kepada Rasulullah Muhammad SAW, sebagai teladan sempurna dalam berakhlak, beribadah, berinteraksi dengan sesama, dan atas syafaatnya yang telah menuntun umat manusia dari jalan kebatilan ke jalan yang benar. Semoga kita semua termasuk dalam golongan yang dituntun Allah SWT dan mendapat syafaat Nabi Muhammad SAW. Aamiin

Penulis sangat menyadari bahwa penulis masih sangat minim ilmu dan pengetahuan, sehingga tanpa adanya peran dan kontribusi dari berbagai pihak yang telah membantu meluangkan waktu dan memberikan sumbangsih pemikiran dalam membimbing penulis, penulis tidak akan mungkin menyelesaikan skripsi ini dengan baik. Selanjutnya penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dan mendukung terselesaikannya skripsi ini.

Ucapan terima kasih ini penulis sampaikan kepada :

1. Prof. Dr. M. Zainuddin, MA selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Hariani, M.Si selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

3. Dr. Fachrul Kurniawan M.MT., IPM, selaku ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Okta Qomaruddin Aziz, M.Kom selaku pembimbing I yang telah meluangkan waktunya untuk memberikan ilmu dan arahnya dalam penyusunan skripsi ini.
5. Dr. Cahyo Crysdiyan, MCS selaku pembimbing II yang telah memberikan ilmu, masukan dan arahan kepada penulis selama penyusunan skripsi ini.
6. Johan Ericka Wahyu Prakasa, M.Kom selaku dosen wali yang selalu memberikan motivasi dan arahan kepada penulis dari awal perkuliahan hingga saat ini.
7. Teman-teman Teknik Informatika Angkatan 2018 yang telah memberikan dukungan serta doa, dan menjadi teman diskusi dalam menuntut ilmu.
8. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan kontribusi dan dukungan selama penyusunan skripsi ini.

Penulis menyadari dalam penyusunan skripsi ini masih terdapat banyak kekurangan, baik dari segi keilmuan, tata bahasa, dan penulisan. Semoga penulisan skripsi ini dapat memberikan manfaat bagi banyak pihak, khususnya terhadap penulis sendiri.

Malang, 22 Juni 2023

Penulis

DAFTAR ISI

| | |
|--|-------------|
| HALAMAN PERSETUJUAN | iii |
| HALAMAN PENGESAHAN | iv |
| HALAMAN MOTTO | vi |
| HALAMAN PERSEMBAHAN | vii |
| KATA PENGANTAR..... | ix |
| DAFTAR ISI..... | xi |
| DAFTAR GAMBAR..... | xii |
| DAFTAR TABEL | xiii |
| ABSTRAK | xiv |
| ABSTRACT | xv |
| مستخلص البحث..... | xvi |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Pernyataan Masalah | 4 |
| 1.3 Tujuan Penelitian | 4 |
| 1.4 Manfaat Penelitian | 4 |
| 1.5 Batasan Masalah..... | 4 |
| BAB II STUDI PUSTAKA | 6 |
| 2.2 Klasifikasi | 6 |
| 2.3 Metode K-Means..... | 7 |
| 2.3.1 Tahapan Metode <i>K-Means</i> | 11 |
| BAB III DESAIN PENELITIAN..... | 14 |
| 3.1 Desain Penelitian..... | 14 |
| 3.2 Pengumpulan Data | 15 |
| 3.3 <i>Preprocessing</i> | 15 |
| 3.4 Desain Sistem..... | 18 |
| BAB IV UJI COBA DAN PEMBAHASAN | 26 |
| 4.1 Skenario Pengujian..... | 26 |
| 4.2 Hasil Uji Coba..... | 28 |
| 4.2.1 Hasil Pengujian Model A..... | 29 |
| 4.2.2 Hasil Pengujian Model B..... | 30 |
| 4.2.3 Hasil Pengujian Model C..... | 31 |
| 4.2.4 Pengujian <i>K-Fold Cross Validation</i> | 32 |
| 4.3 Pembahasan..... | 33 |
| BAB V KESIMPULAN DAN SARAN | 45 |
| 5.1 Kesimpulan | 45 |
| 5.2 Saran..... | 46 |
| DAFTAR PUSTAKA | 47 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1 Tahapan Metode <i>K-Means</i> | 12 |
| Gambar 3. 1 Desain Penelitian..... | 14 |
| Gambar 3. 2 Desain Sistem..... | 18 |
| Gambar 4. 1 Grafik Perbandingan Nilai <i>Precision</i> Setiap Kelas..... | 34 |
| Gambar 4. 2 Grafik Perbandingan Nilai <i>Recall</i> Setiap Kelas | 35 |
| Gambar 4. 3 Perbandingan Nilai Rata-rata <i>Precision</i> dari Ketiga Model..... | 36 |
| Gambar 4. 4 Perbandingan Nilai Rata-rata <i>Recall</i> dari Ketiga Model..... | 37 |
| Gambar 4. 5 Perbandingan Nilai Akurasi setiap Model | 38 |
| Gambar 4. 6 Diagram Perbandingan Performa Setiap Model..... | 38 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2. 1 Penelitian Terkait Mengenai Implementasi Metode <i>K-Means</i> | 8 |
| Tabel 3. 1 Perubahan Nilai Atribut Level | 16 |
| Tabel 3. 2 Nilai <i>Centroid</i> Pertama | 20 |
| Tabel 3. 3 Penentuan <i>Cluster</i> Iterasi Pertama | 21 |
| Tabel 3. 4 Hasil <i>Cluster</i> Iterasi Kedelapan | 23 |
| Tabel 3. 5 Perbandingan Hasil dari Perhitungan <i>Excel</i> dengan Data Aktual..... | 24 |
| Tabel 4. 1 Pembagian Dataset..... | 26 |
| Tabel 4. 2 Jumlah Data Setiap Model | 27 |
| Tabel 4. 3 <i>Confusion Matrix</i> Pengujian Model A | 29 |
| Tabel 4. 4 Nilai <i>Recall</i> dan <i>Precision</i> Setiap Kelas untuk Pengujian Model A.... | 29 |
| Tabel 4. 5 <i>Confusion Matrix</i> Pengujian Model B | 30 |
| Tabel 4. 6 Nilai <i>Recall</i> dan <i>Precision</i> Setiap Kelas untuk Pengujian Model B | 30 |
| Tabel 4. 7 <i>Confusion Matrix</i> Pengujian Model C | 31 |
| Tabel 4. 8 Nilai <i>Recall</i> dan <i>Precision</i> Setiap Kelas untuk Pengujian Model C | 32 |
| Tabel 4. 9 Hasil Akurasi Menggunakan Pengujian <i>K-Fold Cross Validation</i> | 33 |

ABSTRAK

Husna, Hida Muhimmatul. 2023. **Klasifikasi Penyakit Kanker Paru-Paru Berdasarkan Faktor Risiko Menggunakan Metode *K-Means***. Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Okta Qomaruddin Aziz, M.Kom (II) Dr. Cahyo Crysodian, MCS.

Kata Kunci : *Klasifikasi, Kanker Paru, K-Means*

Penyakit kanker paru merupakan penyakit mematikan yang memiliki faktor risiko yang beragam. Faktor risiko yang terkait, diantaranya seperti riwayat merokok, paparan zat berbahaya, dan riwayat keluarga. Banyak orang yang terdiagnosis kanker paru terlambat mencari bantuan dikarenakan tidak menyadari gejala yang dialaminya, karena pada stadium awal sering terjadi tanpa gejala. Berdasarkan permasalahan tersebut, pada penelitian ini dilakukan pengklasifikasian penyakit kanker paru berdasarkan faktor risiko, dengan tujuan dapat membantu dalam diagnosis dini dan perencanaan perawatan yang tepat. Data yang digunakan diperoleh dari sistus data.world, datanya berisikan dataset pasien kanker paru berjumlah 1000 sampel, dan memiliki 22 atribut. Data ini dianalisis dengan menggunakan metode *K-Means*, merupakan metode pengelompokan yang populer dalam analisis data. Data diolah dari langkah mengubah tipe data kategori menjadi numerik, memisahkan data *training* dan data *testing* dengan 3 rasio perbandingan, 90:10, 80:20, dan 70:30, kemudian implementasi metode *K-Means* untuk klasifikasi. Untuk pengujian performa metode dalam melakukan klasifikasi dilakukan pengukuran evaluasi dengan memanfaatkan *confusion matrix*, untuk mendapatkan nilai *accuracy*, *precision*, dan *recall*. Hasil analisis menunjukkan bahwa pada rasio 90:10 dengan data yang digunakan untuk pengujian berjumlah 1000 data, menghasilkan nilai performa yang paling baik. Diperoleh nilai *accuracy* sebesar 73%, *precision* sebesar 73.61%, dan *recall* sebesar 73.51%.

ABSTRACT

Husna, Hida Muhimmatul. 2023. **Classification of Lung Cancer Based on Risk Factors Using the K-Means Method.** Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Okta Qomaruddin Aziz, M.Kom (II) Dr. Cahyo Crysdiyan, MCS.

Lung cancer is a deadly disease with various risk factors. Some of the associated risk factors include smoking history, exposure to harmful substances, and family history. Many people diagnosed with lung cancer seek help late because they are unaware of the symptoms, as early-stage lung cancer often occurs without noticeable symptoms. Based on this problem, this research aims to classify lung cancer based on risk factors to assist in early diagnosis and appropriate treatment planning. The data used was obtained from the data.world dataset, consisting of 1000 samples of lung cancer patients with 22 attributes. The data was analyzed using the K-Means method, which is a popular clustering method in data analysis. The data was processed by converting categorical data into numerical data, separating the data into training and testing sets using three different ratios: 90:10, 80:20, and 70:30, and implementing the K-Means method for classification. The performance of the method in classification was evaluated using the confusion matrix to obtain accuracy, precision, and recall values. The analysis results showed that the 90:10 ratio with a dataset of 1000 samples used for testing yielded the best performance. The accuracy was found to be 73%, precision 73.61%, and recall 73.51%.

Keywords : *Classification, Lung Cancer, K-Means*

مستخلص البحث

الحسنى، هذا مهمة . ٢٠٢٣ . تصنيف سرطان الرئة بناءً على عوامل الخطر باستخدام طريقة *K-Means* رسالة علمية. قسم هندسة المعلومات، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرفون: (I) عكتا قمر الدين عزيز، م. كوم (II) الدكتور كاهيو كريسديان، م. سي. إس.

الكلمات الرئيسية: التصنيف، سرطان الرئة، K-Means

مرض سرطان الرئة هو مرض قاتل يحمل عوامل خطر متنوعة. تشمل هذه العوامل تاريخ التدخين السابق، والتعرض للمواد الضارة، وتاريخ الأمراض في العائلة. العديد من الأشخاص الذين يتم تشخيصهم بسرطان الرئة يتأخرون في طلب المساعدة بسبب عدم إدراكهم للأعراض التي يعانون منها، حيث يحدث ذلك في المراحل المبكرة دون أعراض واضحة. استنادًا إلى هذه المشكلة، تم إجراء هذه الدراسة لتصنيف سرطان الرئة بناءً على العوامل الخطرية، بهدف المساعدة في التشخيص المبكر والتخطيط للعلاج المناسب. تم الحصول على البيانات المستخدمة من منصة data.world ، وتحتوي هذه البيانات على مجموعة من مرضى سرطان الرئة يبلغ عددهم 1000 عينة، وتحتوي على 22 سمة. تم تحليل هذه البيانات باستخدام طريقة K-Means ، وهي طريقة شهيرة لتجميع البيانات. تم معالجة البيانات بتحويل البيانات الفئوية إلى بيانات رقمية، وتقسيم بيانات التدريب والاختبار باستخدام ثلاثة نسب مختلفة: 90:10، 80:20، و 70:30، ثم تم تنفيذ طريقة K-Means للتصنيف. تم قياس أداء الطريقة في التصنيف من خلال استخدام مصفوفة الارتباك (confusion matrix) للحصول على قيم الدقة، الدقة الاستدلالية، والاستدعاء. أظهرت نتائج التحليل أن النسبة 90:10 مع استخدام 1000 عينة للاختبار أعطت أف تظهر نتائج التحليل أن النسبة 90:10 بيانات الاختبار المستخدمة التي تبلغ 1000 عينة أعطت أفضل أداء. تم الحصول على نسبة دقة بلغت 73%، دقة استدلالية بلغت 73.61%، واستدعاء بلغت 73.51%.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu penyakit yang tergolong tidak menular dan menjadi beban kesehatan di seluruh dunia yaitu penyakit kanker (Pangribowo, 2019). Menurut laporan *Global Burden of Cancer* (Globocan) dari *World Health Organization* (WHO) menyatakan bahwa jumlah kematian yang diakibatkan oleh kanker di Indonesia pada tahun 2020 mencapai 234.511 orang, berdasarkan jenisnya kasus kematian pada kanker paru menjadi tertinggi, jumlahnya mencapai 30.843 orang atau 13,2% dari semua kematian akibat kanker (databoks.katadata.co.id, 2022). Kanker paru merupakan penyakit yang memiliki ciri khas adanya pertumbuhan sel ganas pada jaringan paru-paru. Apabila tidak segera dirawat pertumbuhan sel ini bisa tidak terkontrol dan penyebarannya bisa ke jaringan dan organ sekitar (Amalita & Dewi, 2021).

Kanker paru-paru banyak terjadi pada orang yang melakukan kebiasaan dan aktivitas yang dapat menyebabkan masalah paru-paru seperti merokok, yang merupakan salah satu penyebab utama kanker di Indonesia (Prasetio & Susanti, 2019). Namun, selain kebiasaan negatif tersebut, beberapa makanan juga bisa memicu terjadinya kanker paru, seperti makanan tinggi lemak jenuh dan karbohidrat olahan (Risnawati et al., 2020). Makanan sangat memiliki peranan penting pada kesehatan sel pasien kanker paru, oleh sebab itu pasien perlu memperhatikan makanan yang dikonsumsi. Sebagaimana firman Allah dalam Surat Abasa ayat 24 :

فَلْيَنْظُرِ الْإِنْسَانُ إِلَى طَعَامِهِ

“Maka hendaklah manusia itu memperhatikan makanannya” (Q.S. Abasa : 24).

Perhatian Allah dalam surat diatas mengandung hikmah dan pelajaran yang berharga. Setiap orang dalam mengkonsumsi makanan memang harus selektif supaya bahan makanan yang dikonsumsi tidak berbahaya bagi tubuh, higienis, dan bergizi. Untuk mengurangi risiko terkena kanker paru, pasien disarankan menyesuaikan asupan makanannya dengan memperbanyak jumlah sayur dan buah, nutrisi yang terkandung dalam sayuran dan buah-buahan secara positif dapat mencegah pembentukan sel kanker, karena adanya vitamin, mineral, dan serat yang terkandung di dalamnya (Barta et al., 2019).

Sebab kanker paru-paru merupakan penyakit yang mematikan karena sulit dideteksi sebelum menjadi penyakit yang serius. Banyak orang yang terdiagnosis kanker paru terlambat mencari bantuan dikarenakan tidak menyadari gejala yang dialaminya, karena pada stadium awal sering terjadi tanpa gejala (Febriani & Furqon, 2020). Sekitar 85% kanker paru-paru baru bisa terdeteksi pada stadium lanjut (Li et al., 2018). Kanker paru juga tidak dapat dilihat tanpa CT *scan* atau *rontgen* dada. Sebagian besar kasus kanker paru menyerang orang berusia antara 35 dan 75 tahun, dengan insiden tertinggi antara 55 dan 65 tahun (Aliyah et al., 2016).

Berdasarkan permasalahan tersebut, jika faktor penyebab kanker paru-paru tidak ditanggapi dengan serius maka jumlah penderita kanker paru-paru pasti akan meningkat seiring berjalannya waktu, oleh karena itu diperlukan suatu pendekatan

untuk mengidentifikasi status pasien kanker paru, apakah tergolong rendah, sedang atau tinggi. Pada penelitian yang telah dilakukan oleh Abrianto (2018) menggunakan metode *k-means* dalam mengkategorikan pasien penyakit liver, metode yang digunakan pada sistem berjalan dengan baik dimana ditunjukkan pada pengujian hasil akhir metode dengan tingkat kesesuaian 86% untuk membandingkan perhitungan sistem dengan perhitungan manual, untuk perbandingan hasil perhitungan sistem dengan hasil pakar tingkat kesesuaiannya 66%.

Pengujian nilai akurasi hasil klasifikasi bertujuan untuk mendeteksi kesalahan klasifikasi, sehingga dapat diketahui presentase nilai akurasi dari hasil klasifikasi. (Bashit et al., 2019). Uji akurasi adalah suatu hal penting sebagai penilaian kualitas hasil klasifikasi. Uji akurasi dilakukan untuk mengetahui tingkat kesesuaian hasil klasifikasi dengan data ukuran sebenarnya. Jika dalam dunia medis, nilai presentase dari uji akurasi penting dijadikan sebagai acuan karena dapat diketahui tingkat kepercayaan terhadap pemakaian hasil klasifikasi dengan kebutuhan yang diperlukan.

Penelitian ini akan menggunakan metode *k-means* untuk klasifikasi penyakit kanker paru-paru berdasarkan faktor risiko, dan digunakan tabel *confusion matrix* sebagai uji akurasi hasil klasifikasi. Metode *k-means* menjadi salah satu metode pada data mining yang sederhana dan populer yang digunakan untuk memecahkan masalah pengelompokan data (Silitonga & Morina, 2017). Diharapkan hasil yang telah didapat, nantinya bermanfaat untuk dunia medis, seperti dapat melakukan deteksi dini kanker paru pada pasien.

1.2 Pernyataan Masalah

Seberapa besar tingkat akurasi dari penggunaan metode *k-means* dalam mengklasifikasi penyakit kanker paru berdasarkan faktor risiko?

1.3 Tujuan Penelitian

Mengukur tingkat akurasi metode *k-means* dalam mengklasifikasi penyakit kanker paru berdasarkan faktor risiko.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang sudah dijelaskan, maka diharapkan skripsi ini memiliki manfaat sebagai berikut :

1. Dapat membantu pihak Institusi Kesehatan untuk melakukan deteksi dini kanker paru pada pasien yang memiliki faktor penyebab kanker paru.
2. Dapat membantu pasien kanker paru untuk mengetahui status penyakitnya apakah tergolong rendah, sedang atau tinggi.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data rekam medis pasien kanker paru yang berjumlah 1000 sampel, dan terdapat 22 atribut yang terdiri dari *gender, air pollution, alcohol use, dust allergy, occupational hazards, 'genetic risk, chronic lung disease, balanced diet, obesity, smoking, passive smoker, chest pain, coughing of blood, fatigue, weight loss, 'shortness of breath, wheezing, swallowing difficulty, clubbing of finger nails, frequent cold, dry cough, snoring.*

2. Dataset yang digunakan adalah data sekunder yang diperoleh dari website :

<https://data.world/cancerdatahp/lung-cancer-data>

BAB II

STUDI PUSTAKA

2.1 Kanker Paru

Kanker paru adalah suatu keganasan pada paru-paru yang disebabkan oleh perubahan genetika pada sel epitel saluran nafas, sehingga terjadi proliferasi sel (siklus pembelahan sel) yang tak terkendali (Buana & Harahap, 2022). Keganasan tersebut bisa bermula dari organ paru itu sendiri (primer) atau dari luar paru (metastasis). Individu yang mengidapnya sering tidak menyadari bahwa penyakit itu ada, karena gejala penyakit ini sering tidak dikenali pada tahap tertentu. Pada stadium awal, kanker paru tidak begitu menimbulkan gejala yang menonjol, namun dari waktu ke waktu gejala umum yang bisa dilihat adalah batuk yang semakin memburuk dan tidak pernah reda, sesak nafas, merasa nyeri secara konstan di bagian dada, batuk berdarah, sering terkena infeksi paru, merasa letih setiap saat, dan mengalami penurunan berat badan tanpa sebab yang jelas (Toloza et al., 2003).

Sebagian besar kanker merupakan hasil dari banyak faktor risiko. Faktor risiko merupakan sesuatu yang meningkatkan risiko terkena kanker paru, bisa karena perilaku, kondisi, atau substansi. Beberapa faktor risiko kanker paru yaitu umur, jenis kelamin, kebiasaan merokok, perokok pasif, paparan zat karsinogen, polusi udara, dan pengaruh genetik (Prabowo et al., 2022).

2.2 Klasifikasi

Menurut KBBI (Kamus Besar Bahasa Indonesia) klasifikasi merupakan pengelompokan suatu sistem ke dalam kelompok-kelompok menurut ketentuan-

ketentuan yang telah ditetapkan. Secara harfiah juga dapat dikatakan bahwa klasifikasi merupakan pembagian sesuatu menurut kelas-kelas (Lestari, 2016). Klasifikasi adalah proses mengelompokkan data berdasarkan kesamaan maupun perbedaan (Subrata, 2019). Oleh karena itu, dapat disimpulkan bahwa klasifikasi kanker paru merupakan pengelompokan status pasien kanker paru berdasarkan kelas yang sudah ditetapkan. Tujuan dari pengelompokan ini yaitu untuk memberikan status pada pasien, apakah kanker paru yang dialaminya tergolong rendah, sedang atau tinggi.

Pada penelitian yang dilakukan oleh Achmad Ridwan tentang klasifikasi gejala dini penyakit Diabetes Mellitus menggunakan metode *Naive Bayes* pada dataset pasien dengan gejala yang dapat menyebabkan diabetes, 520 dataset dengan 17 variabel digunakan sebagai parameter dan label kelas dengan nilai positif dan negatif. Peneliti mengevaluasi model klasifikasi menggunakan *confusion matrix*, diperoleh nilai akurasi prediksi data sebesar 90.20% (Ridwan, 2020).

Pada penelitian Athalla *et al.* juga melakukan pengklasifikasian pada penyakit kanker payudara menggunakan metode *K-NN*, dataset yang digunakan memiliki jumlah sampel sebanyak 569 dengan 30 variabel dan label kelas yang dimiliki bernilai jinak dan ganas, perhitungan jarak yang digunakan yaitu jarak *Minkowski* yang telah ditentukan dari kedua jarak *Euclidean* dan jarak *Manhattan*, hasil akurasi yang didapatkan sebesar 93% (Athalla et al., 2018).

2.3 Metode K-Means

K-means adalah metode *clustering* yang dapat mengelompokkan data dengan karakteristik yang sama dikelompokkan dalam satu kelompok atau *cluster*

dan data dengan karakteristik berbeda dikelompokkan dalam *cluster* lain (Metisen & Sari, 2015). Ukuran dari kemiripan data diukur dengan seberapa dekat data tersebut dengan rata-rata *cluster*. Metode *k-means* merupakan salah satu algoritma dalam *data mining* yang dapat digunakan untuk melakukan pengelompokan suatu data.

Cara kerja dari metode *k-means* yaitu, data-data yang menjadi inputan akan dikelompokkan sendiri oleh sistem, tanpa mengetahui terlebih dahulu target kelasnya. Inputan yang diterima oleh sistem berupa data dan *k* buah *cluster*, simbol *k* merupakan jumlah *cluster* yang diinginkan. Metode *k-means* nantinya akan mengelompokkan data ke dalam *k* buah *cluster* tersebut, dan untuk bisa mengelompokkan data, metode *k-means* bekerja secara iteratif dalam menghitung jarak dari setiap data ke titik *centroid*. Proses tersebut berhenti ketika nilai *centroid* telah stabil atau berarti sudah tidak terjadi perubahan nilai, dan juga anggota dari setiap *cluster* tidak berpindah-pindah lagi.

Penelitian-penelitian terkait mengenai metode *k-means* akan dijelaskan dalam sebuah tabel, yaitu sebagai berikut :

Tabel 2. 1 Penelitian Terkait Mengenai Implementasi Metode *K-Means*

| No. | Peneliti | Metode | Hasil |
|-----|---|----------------|---|
| 1. | Rozzi Kesuma Dinata, Safwandi, Novia Hasdyna, Nur Azizah, (2020) Analisis Metode <i>K-Means</i> pada Data Sepeda Motor | <i>K-Means</i> | Menerapkan <i>k-means</i> dalam pengelompokan data sepeda motor. Hasil dari penelitian ini berupa <i>clustering</i> yang memisahkan data menjadi 3 <i>cluster</i> , yaitu murah, standard, dan mahal. Hasil analisis performansi metode <i>k-means</i> dari 15 pengujian dari setiap uji coba yang dilakukan, diperoleh nilai rata-rata <i>presicion</i> sebesar 76%, nilai <i>recall</i> |

| No. | Peneliti | Metode | Hasil |
|-----|---|--------------------|---|
| | | | sebesar 76% dan <i>accuracy</i> sebesar 81%. |
| 2. | Putra May Chandra Abrianto, (2018) Penerapan Metode <i>K-Means Clustering</i> untuk Pengelompokan Pasien Penyakit Liver | <i>K-Means</i> | Dengan metode <i>K-Means Clustering</i> , hasil dikelompokkan dalam 3 kategori yaitu tinggi, normal dan rendah. Metode <i>K-Means Clustering</i> yang diterapkan dalam sistem berjalan dengan tingkat kesesuaian 66% untuk perbandingan hasil perhitungan sistem dengan hasil pakar. |
| 3. | Arifin dan Syalwah, (2020) Memprediksi keefektifan <i>immunotherapy</i> pada penyakit kutil dengan metode <i>Naive Bayes</i> | <i>Naive Bayes</i> | Menggunakan dataset tahun 2018 dengan 90 dataset dan 8 atribut yang didapat dari <i>UCI Repository</i> . Pada penelitian prediksi dibagi menjadi dua ke dalam dua kelompok, berhasil dan gagal. Pemisahan data latih dan data uji menggunakan metode <i>cross-validation</i> dengan menetapkan nilai <i>k</i> sama dengan 10, dan menggunakan <i>confusion matrix</i> untuk menghasilkan akurasi sebagai performa metode dengan nilai akurasi sebesar 81,11%. |
| 4. | Chandra Purnamaningsih, Ristu Saptono, Abdul Aziz (2016) Penentuan Penjurusan Siswa SMA menggunakan Metode <i>K-Means Clustering</i> | <i>K-Means</i> | Berdasarkan hasil penelitian, pengujian terbaik pada proses <i>clustering</i> IPA dengan hasil akurasi 0.905882, <i>sensitivity</i> 0.876923, <i>specificity</i> 0.714285. Sedangkan pengujian terbaik juga pada <i>clustering</i> IPS, didapatkan akurasi 0.905882, <i>recall</i> 0.714285. Hasil perbandingan <i>clustering</i> terbaik pada <i>clustering</i> IPA dengan <i>clustering</i> IPS menunjukkan bahwa tidak ada siswa yang diterima di dua jurusan IPA/IPS atau siswa ditolak di keduanya. Hal ini karena proses <i>clustering</i> mempertimbangkan kriteria nilai akademik, IQ dan minat siswa yang dapat mempengaruhi proses hasil <i>clustering</i> menggunakan <i>k-means</i> . |

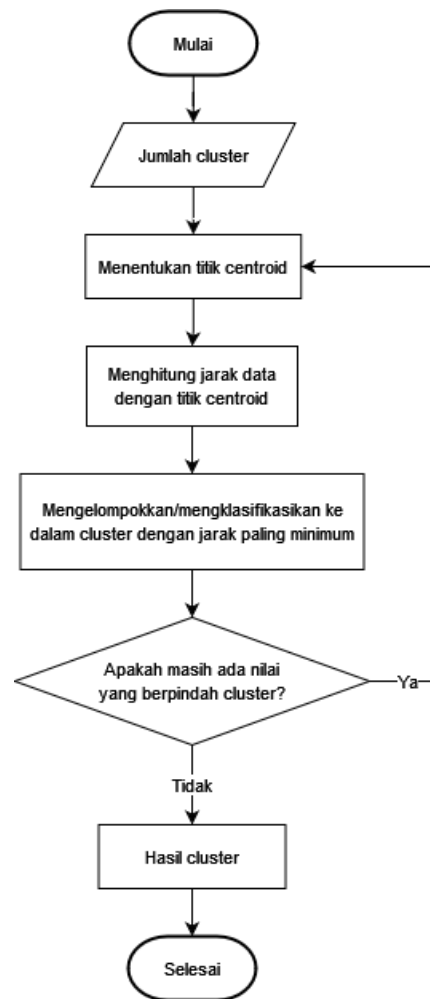
| No. | Peneliti | Metode | Hasil |
|-----|---|---|--|
| 5. | Abu Salam, Diyan Adiatma, Junta Zeniarja (2020) Implementasi Algoritma <i>K-Means</i> dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS | <i>K-Means</i> | Dari hasil proses cluster dihasilkan 154 data pendaftar yang direkomendasikan mendapatkan beasiswa sedangkan 287 data lainnya tidak mendapatkan beasiswa, dari hasil tersebut dilakukan perbandingan dengan hasil penerimaan beasiswa PPA pada periode penerimaan tahun 2016. Dari penggunaan tabel <i>confusion matrix</i> didapatkan nilai <i>precision</i> 73%, <i>recall</i> 73%, dan akurasi 82%. |
| 6. | Waskito Wahyu Pribadi, Amak Yunus, Anggri Sartika Wiguna (2022) Perbandingan Metode K-Means Euclidean Distance dan Manhattan Distance pada Penentuan Zonasi Covid-19 di Kabupaten Malang | <i>K-Means Euclidean Distance</i> , dan <i>Manhattan Distance</i> | Hasil dari perbandingan perhitungan jarak <i>Euclidean</i> dan <i>Manhattan Distance</i> menggunakan <i>Silhouette Coefficient</i> menunjukkan bahwa <i>Euclidean Distance</i> lebih unggul dibandingkan <i>Manhattan Distance</i> dengan hasil perbandingan $0.71 > 0.64$. Sedangkan untuk Hasil dari perhitungan Standar Deviasi menunjukkan bahwa <i>Euclidean Distance</i> lebih unggul dibandingkan <i>Manhattan Distance</i> dengan hasil perbandingan $0.46466002 < 0.4961977$. |
| 7. | Isaias F.X.F.A. Felnditi, Desi Arisandi, Tri Sutrisno, (2020) | <i>K-Means</i> | Menerapkan metode <i>k-means</i> untuk perancangan sistem rekomendasi penjurusan pada sekolah menengah jurusan Santo Paulus. Pada penelitian ini digunakan 200 data siswa, dan diperoleh hasil akurasi sebesar 62.5% dari data minat asli siswa. |
| 8. | Penelitian Penulis Klasifikasi Penyakit Kanker Paru-Paru Berdasarkan Faktor Risiko Menggunakan Metode <i>K- Means</i> | <i>K-Means</i> | Digunakan dataset pasien kanker paru sebanyak 1000 sampel dan memiliki 22 variabel yang merupakan faktor risiko. Faktor risiko dapat mencakup variabel seperti riwayat keluarga, perokok, perokok pasif, dan faktor lainnya. Dilakukan klasifikasi menggunakan <i>k-means</i> untuk mengidentifikasi kelompok-kelompok dengan |

| No. | Peneliti | Metode | Hasil |
|-----|----------|--------|---|
| | | | menemukan pusat <i>cluster</i> yang mewakili setiap kelompok. Karena pada dataset sudah terdapat label kelas yang terdiri dari kelompok <i>high</i> , <i>medium</i> , dan <i>low</i> maka identifikasi kelompok ini dilakukan untuk membedakan tingkat risiko kanker paru yang diderita pasien. |

Pada Tabel 2.1 telah dijelaskan mengenai penelitian-penelitian sebelumnya yang menerapkan metode *k-means*. Perbedaan penelitian ini dengan penelitian-penelitian sebelumnya yaitu terletak pada studi kasusnya, pada penelitian ini akan membangun sebuah sistem dalam mengklasifikasi penyakit kanker paru berdasarkan faktor risiko, dan digunakan *confusion matrix* untuk menghitung akurasi dari performa metode *k-means*.

2.3.1 Tahapan Metode *K-Means*

Pada prosesnya, metode *k-means* membutuhkan beberapa parameter input dan beberapa jumlah *cluster*, kemudian data dibagi ke dalam *cluster – cluster* hingga tercapai tingkat kesamaan data dalam *cluster* tersebut. Berikut adalah langkah-langkah dari metode *k-means*, dapat dilihat pada Gambar 2.1.



Gambar 2. 1 Tahapan Metode *K-Means*

Penjelasan dari setiap tahap metode *k-means* dijelaskan sebagai berikut (Agusta, 2007) :

1. Ditentukannya jumlah *cluster* yang akan dibentuk, biasanya diinisialisasikan oleh variabel k .
2. Menentukan titik *centroid* atau titik pusat *cluster*. Pada awal iterasi, untuk penentuan titik *centroid* diambil secara random dari data yang digunakan. Pada iterasi selanjutnya *titik centroid* ditentukan dengan menggunakan data rata-rata dari setiap *cluster*.

3. Menghitung jarak dari setiap data ke titik *centroid*. Digunakan rumus *euclidean distance* untuk perhitungan jaraknya. Rumus *euclidean distance* adalah sebagai berikut :

$$D(i, j) = \sqrt{\sum_{i=1}^{n_k} (x_{ki} - c_{kj})^2} \quad (2.1)$$

Keterangan :

- (i, j) = Jarak data ke-i ke *centroid* ke-j
 x_{ki} = Data ke-i dalam *cluster* pada atribut data ke-k
 c_{kj} = *Centroid* ke-j pada atribut data ke-k
 n_k = Banyaknya data pada atribut data ke-k

4. Kelompokkan data ke dalam *cluster* sehingga semua data paling dekat dengan setiap *centroid*. Jarak paling dekat antara satu data dengan *cluster* tertentu menentukan suatu data termasuk dalam anggota *cluster* yang mana.
5. Nilai *centroid* diperoleh dari rata-rata semua data yang terdapat dalam setiap *cluster* dengan keanggotaan yang terbaru. Dalam hal ini rumus yang digunakan adalah sebagai berikut :

$$C_{ki} = \frac{1}{n_i} \sum_{q=1}^{n_{ki}} x_{kq} \quad (2.2)$$

Keterangan :

- C_{ki} = Titik *centroid* pada *cluster* ke-i dan atribut ke-k
 x_{kq} = Data ke-q pada *cluster* ke-i dan atribut ke-k
 n_{ki} = Banyaknya data pada *cluster* ke-i dan atribut ke-k

6. Mengulangi langkah 2 sampai 5 hingga tidak ada perubahan atau perpindahan data di setiap *cluster*. Data yang tidak berubah menunjukkan bahwa perulangan telah berakhir dan pengelompokan selesai.

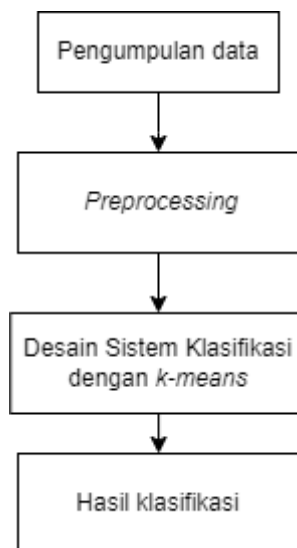
BAB III

DESAIN PENELITIAN

Pada bab ini akan dibahas mengenai perancangan sistem. Mencakup desain penelitian, desain sistem, dan proses perhitungan metode *k-means* yang diterapkan untuk klasifikasi pada dataset pasien kanker paru berdasarkan faktor risiko.

3.1 Desain Penelitian

Desain penelitian merupakan rangkaian tahapan yang digunakan sebagai pedoman dalam alur jalannya penelitian, sehingga penelitian dilakukan lebih terarah. Desain penelitian dapat dilihat pada Gambar 3.1 sebagai berikut.



Gambar 3. 1 Desain Penelitian

Tahapan dalam penelitian ini diawali dengan pengumpulan data berupa data pasien kanker paru yang memiliki beberapa faktor risiko kanker paru, pada proses pengumpulan data ini memanfaatkan data pasien yang sudah disediakan oleh data.world. Selanjutnya pada tahap *preprocessing*, data diolah hingga menjadi data

bersih yang dapat digunakan pada proses selanjutnya. Kemudian model klasifikasi dengan menerapkan metode *K-Means* dibangun untuk mendapatkan *output* atau hasil akhir berupa level penyakit pasien kanker paru, apakah tergolong tinggi, sedang atau rendah. Selanjutnya dilakukan pengujian terhadap sistem yang dibangun dengan mengevaluasi kinerja metode *k-means*.

3.2 Pengumpulan Data

Bentuk dataset yang diperoleh adalah berupa data sekunder, data yang digunakan yaitu data rekam medis pasien kanker paru. Sumber dataset berasal dari website <https://data.world/cancerdatahp/lung-cancer-data>, maka data ini bersifat *public*. Menurut rekam data, terdapat 22 atribut yang menjadi penyebab pasien terkena kanker paru, gambaran mengenai dataset yang digunakan sebagai contoh dapat dilihat pada Tabel 3.2, dari 22 atribut tersebut setiap pasien di rekam berdasarkan level dari tingkat yang rendah sampai level tinggi, yaitu mulai level 1 sampai dengan level 9. Pada dataset juga terdapat satu kolom yang menyatakan sebagai label kelas dataset, yaitu kolom level yang terdiri dari kelas *high* (riwayat pasien kanker paru berstatus tinggi), *medium* (riwayat pasien kanker paru berstatus sedang), dan *low* (riwayat pasien kanker paru berstatus rendah). Dataset disimpan dalam bentuk format *excel*.

3.3 Preprocessing

Tahap *preprocessing* adalah tahapan dilakukannya pemeriksaan pada data yang akan digunakan, tujuannya untuk memaksimalkan hasil. Tahapan *preprocessing* disesuaikan dengan kebutuhan penelitian yang dilakukan. Pada penelitian ini

preprocessing yang dibutuhkan yaitu mengubah tipe data kategori menjadi numerik. Tahap tersebut merupakan *preprocessing* yang mengonversi tipe data pada atribut dataset yang semula berupa data kategori menjadi data numerik, dengan tujuan untuk mempermudah saat pemrosesan data.

Atribut atau variabel dalam dataset pasien kanker paru yang hanya memiliki tipe data kategori adalah atribut level, dengan nilai *high* (riwayat pasien kanker paru berstatus tinggi), *medium* (riwayat pasien kanker paru berstatus sedang), dan *low* (riwayat pasien kanker paru berstatus rendah). Dari nilai tersebut nilai *high*, *medium*, dan *low* diubah menjadi nilai numerik 0, 2, dan 1. Perubahan nilai tipe data pada atribut level dapat dilihat dalam Tabel 3.1.

Tabel 3. 1 Perubahan Nilai Atribut Level

| Data Awal Atribut Level | Hasil Perubahan Data Atribut Level |
|-------------------------|------------------------------------|
| <i>Low</i> | 1 |
| <i>Medium</i> | 2 |
| <i>High</i> | 0 |

Berdasarkan dataset yang sudah didapat maka tidak diterapkan *preprocessing* tahap data *scaling*, karena data numerik pada dataset sudah memiliki rentang nilai yang sama, contoh data dapat dilihat pada Tabel 3.2. Umumnya data *scaling* dilakukan dengan tujuan untuk mengubah skala data agar memiliki rentang nilai yang seragam atau normal, sehingga memudahkan dalam proses perhitungan dan memperbaiki performa model sistem.

Tabel 3. 2 Data rekam medis pasien kanker paru

| NO | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------|
| 1 | 33 | 2 | 4 | 5 | 4 | 3 | 2 | 2 | 4 | 3 | 2 | 2 | 4 | 3 | 4 | 2 | 2 | 3 | 1 | 2 | 3 | 4 | Low |
| 2 | 37 | 7 | 7 | 7 | 7 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 8 | 4 | 2 | 3 | 1 | 4 | 5 | 6 | 7 | 5 | High |
| 3 | 52 | 2 | 4 | 5 | 4 | 3 | 2 | 2 | 4 | 3 | 2 | 2 | 4 | 3 | 4 | 2 | 2 | 3 | 1 | 2 | 3 | 4 | Low |
| 4 | 35 | 4 | 5 | 6 | 5 | 6 | 5 | 5 | 5 | 6 | 6 | 6 | 5 | 1 | 4 | 3 | 2 | 4 | 6 | 2 | 4 | 1 | Medium |
| 5 | 46 | 2 | 3 | 4 | 2 | 4 | 3 | 3 | 3 | 2 | 3 | 4 | 4 | 1 | 2 | 4 | 6 | 5 | 4 | 2 | 1 | 5 | Medium |
| 6 | 28 | 3 | 1 | 4 | 3 | 2 | 3 | 4 | 3 | 1 | 4 | 3 | 1 | 3 | 2 | 2 | 4 | 2 | 2 | 3 | 4 | 3 | Low |
| 7 | 34 | 6 | 7 | 7 | 7 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 8 | 4 | 2 | 3 | 1 | 4 | 5 | 6 | 7 | 5 | High |
| 8 | 62 | 6 | 8 | 7 | 7 | 7 | 6 | 7 | 7 | 8 | 7 | 7 | 9 | 3 | 2 | 4 | 1 | 4 | 2 | 4 | 2 | 3 | High |
| 9 | 32 | 1 | 6 | 7 | 8 | 7 | 6 | 7 | 7 | 3 | 4 | 8 | 7 | 3 | 2 | 6 | 4 | 2 | 3 | 1 | 2 | 1 | Medium |
| 10 | 54 | 6 | 7 | 7 | 7 | 7 | 6 | 7 | 7 | 7 | 8 | 7 | 7 | 5 | 3 | 2 | 7 | 8 | 2 | 4 | 5 | 3 | High |

Keterangan :

A : Age

B : Air Pollution

C : Alcohol use

D : Dust Allergy

E : OccuPational Hazards

F : Genetic Risk

G : Chronic Lung Disease

H : Balanced Diet

I : Obesity

J : Smoking

K : Passive Smoker

L : Chest Pain

M : Coughing of Blood

N : Fatigue

O : Weight Loss

P : Shortness of Breath

Q : Wheezing

R : Swallowing Difficulty

S : Clubbing of FingerNails

T : Frequent Cold

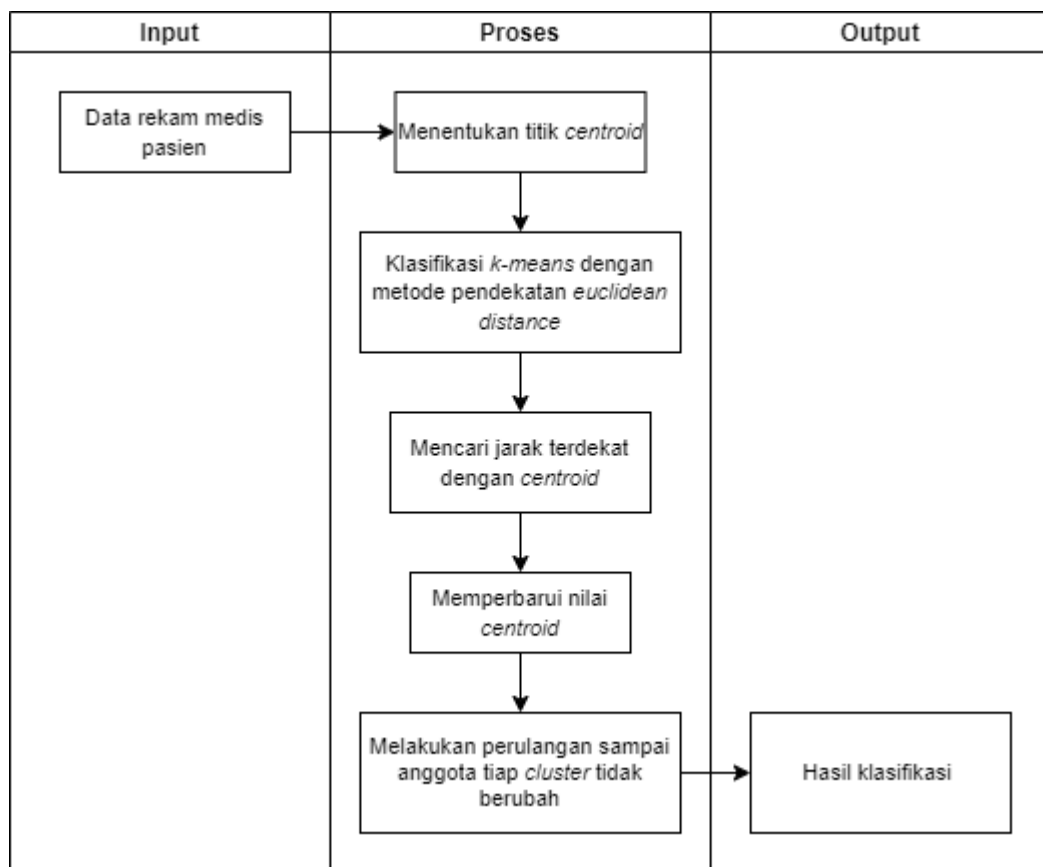
U : Dry Cough

V : Snoring

W : Level

3.4 Desain Sistem

Desain sistem ini merupakan bagian yang menjelaskan dari alur jalannya sistem, terdiri dari tiga tahap, yaitu : *input*, proses, dan *output*. Gambaran alur ini dapat dilihat pada Gambar 3.3 yang merupakan desain sistem klasifikasi menggunakan metode *k-means*.



Gambar 3. 2 Desain Sistem

Berdasarkan desain sistem yang terlihat pada Gambar 3.3, langkah awal sistem dimulai dengan memasukkan dataset pasien kanker paru yang sudah dilakukan *preprocessing*, sebagai input. Langkah berikutnya adalah menentukan titik *centroid*, dalam pengujian ini dataset akan dikelompokkan menjadi 3 *cluster* sesuai dengan jumlah kelas pada dataset pasien, sehingga diperlukan 3 titik

centroid. Langkah selanjutnya adalah melakukan perhitungan jarak antara setiap data dengan pusat *cluster* atau titik *centroid* menggunakan rumus *Euclidian distance*. Setelah itu, data akan dikelompokkan berdasarkan nilai yang diperoleh sesuai dengan *cluster* yang telah dihitung. Selanjutnya menghitung jarak terdekat kembali menggunakan *centroid* baru. Setelah itu dilakukan perulangan pada langkah perhitungan mencari jarak terdekat, sampai anggota tiap *cluster* tidak ada yang berubah. Setelah itu akan didapatkan *output* hasil klasifikasi berupa status kanker paru yang diderita pasien, yaitu terdiri dari level *high*, *medium*, dan *low*. Penjelasan dari masing-masing tahapan tersebut dijelaskan sebagai berikut :

1. Input yang digunakan adalah dataset pasien kanker paru yang sudah dilakukan *preprocessing*, untuk proses *preprocessing* seperti yang sudah dijelaskan pada sub bab *preprocessing*. Kumpulan dataset tersebut tersedia dalam format *excel* yang terdiri dari 1000 sampel dan 22 atribut yang mencakup *gender*, *air pollution*, *alcohol use*, *dust allergy*, *occupational hazards*, *genetic risk*, *chronic lung disease*, *balanced diet*, *obesity*, *smoking*, *passive smoker*, *chest pain*, *coughing of blood*, *fatigue*, *weight loss*, *shortness of breath*, *wheezing*, *swallowing difficulty*, *clubbing of finger nails*, *frequent cold*, *dry cough*, *snoring*, dan atribut level sebagai label kelas dari dataset yang terdiri dari kelas *high* (riwayat pasien kanker paru berstatus tinggi), *medium* (riwayat pasien kanker paru berstatus sedang), dan *low* (riwayat pasien kanker paru berstatus rendah).
2. Selanjutnya yaitu menentukan nilai awal *centroid*, karena jumlah *cluster* yang ditentukan adalah 3 sesuai dengan jumlah kelas di dataset pasien yaitu terdapat

kelas *high*, *medium*, dan *low* maka diperlukan 3 titik *centroid*. Menurut aturan metode *k-means*, penentuan titik awal *centroid* dilakukan secara random. Berikut adalah contoh titik *centroid* pertama yang diambil secara random, seperti yang terlihat pada Tabel 3.2 :

Tabel 3. 2 Nilai *Centroid* Pertama

| | Age | Smoking | Dry cough | Genetic risk | Dust allergy | | Passive smoker | Keterangan |
|----|-----|---------|-----------|--------------|--------------|------|----------------|-------------|
| C1 | 44 | 7 | 5 | 7 | 7 | | 8 | Data ke-100 |
| C2 | 38 | 2 | 4 | 2 | 3 | | 5 | Data ke-500 |
| C3 | 48 | 1 | 4 | 7 | 7 | | 2 | Data ke-700 |

Keterangan : C1 = High
C2 = Low
C3 = Medium

- Setelah itu menghitung jarak antara data dengan titik *centroid*, untuk menghitung jarak dari setiap data ke titik *centroid* dapat digunakan *Euclidean distance*, atau *Manhattan distance*. Pada penelitian ini perhitungan yang dilakukan untuk klasifikasi menggunakan persamaan 2.1 yaitu *Euclidean distance*.
- Selanjutnya mengelompokkan seluruh data ke dalam *cluster*. Dilakukan perhitungan *centroid* atau rata-rata dari data yang ada pada masing-masing *cluster* menggunakan *Euclidean distance*.

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (3.1)$$

Keterangan :
 d = jarak dengan *centroid*
 i = banyaknya atribut pada data pasien
 x dan y = nilai dari atribut pasien
 s dan t = nilai titik *centroid*

5. Setelah dilakukan perhitungan tersebut, data yang memiliki jarak terdekat dengan suatu *cluster* akan menjadi anggota dari *cluster* tersebut.
6. Pada pelatihan awal perhitungan diproses menggunakan *Microsoft Excel*. Hasil dari iterasi pertama perhitungan jarak antara data dengan titik *centroid* setiap *cluster* dapat dilihat pada Tabel 3.3, dari 1000 sampel dataset pasien, peneliti memberikan 20 data sebagai contoh.

Tabel 3. 3 Penentuan *Cluster* Iterasi Pertama

| Data ke- i | C1 | C2 | C3 | Terdekat | Cluster |
|------------|---------|--------|---------|-----------|---------|
| 1 | 18 | 17.146 | 18.5472 | 17.146428 | C2 |
| 2 | 11.4018 | 17.146 | 18.9737 | 17.146428 | C2 |
| 3 | 9 | 19 | 16.3401 | 9 | C1 |
| 4 | 18 | 17.146 | 18.5472 | 17.146428 | C2 |
| 5 | 0 | 19.391 | 17.1464 | 0 | C1 |
| 6 | 22.2036 | 32.726 | 21.9317 | 21.931712 | C3 |
| 7 | 13.4164 | 17.493 | 20.8327 | 13.416408 | C1 |
| 8 | 13.4164 | 17.493 | 20.8327 | 13.416408 | C1 |
| 9 | 12.4097 | 21.119 | 18.868 | 12.409674 | C1 |
| 10 | 22.5389 | 22.361 | 28.9482 | 22.36068 | C2 |
| 11 | 18 | 22.583 | 14.1421 | 14.142136 | C3 |
| 12 | 20.025 | 29.547 | 21.4243 | 20.024984 | C1 |
| 13 | 17.5499 | 21.071 | 22.7596 | 17.549929 | C1 |
| 14 | 12.4097 | 21.119 | 18.868 | 12.409674 | C1 |
| 15 | 12.8452 | 18.083 | 19.7737 | 12.845233 | C1 |
| 16 | 13.9284 | 16.248 | 20.4939 | 13.928388 | C1 |
| 17 | 14.7309 | 18.248 | 15.7797 | 14.73092 | C1 |
| 18 | 22.5389 | 22.361 | 28.9482 | 22.36068 | C2 |
| 19 | 18 | 22.583 | 14.1421 | 14.142136 | C3 |
| 20 | 20.025 | 29.547 | 21.4243 | 20.024984 | C1 |

Berdasarkan Tabel 3.4 hasil yang didapat dari perhitungan iterasi pertama. C1 menyatakan *cluster* pertama, C2 menyatakan *cluster* kedua, dan C3 menyatakan *cluster* ketiga. Data dengan jarak minimal akan dimasukkan pada kolom terdekat, kemudian data tersebut yang memiliki jarak terdekat dengan suatu *cluster* akan menjadi anggota *cluster* tersebut. Pada proses iterasi pertama seperti yang sudah dijelaskan sebelumnya, nilai *centroid* yang digunakan merupakan nilai asli dari dataset pasien yang mana diambil secara random.

7. Langkah selanjutnya adalah menentukan titik *centroid* baru, dilakukan perhitungan menggunakan persamaan 2.2 yaitu dengan mengambil rata-rata dari semua anggota *cluster* dengan menjumlahkan nilai-nilai *cluster* dan membaginya dengan jumlah anggota.
8. Setelah mendapatkan pusat *cluster* yang baru, langkah selanjutnya adalah kembali ke langkah menghitung jarak setiap data ke pusat *cluster*.
9. Proses iterasi akan terus berlanjut sampai titik *centroid cluster* tidak berubah lagi dan nilai yang dihasilkan sudah tidak ada perubahan, dan juga anggota *cluster* yang dihasilkan sudah tidak berpindah-pindah.
10. Pada perhitungan menggunakan langkah-langkah diatas, didapatkan bahwa perhitungan berhenti pada iterasi ke- 8. Berdasarkan hasil iterasi ke- 8 nilai dari C1, C2, dan C3 memiliki data yang sama dengan iterasi sebelumnya. Titik pusat *cluster* yang dihasilkan sudah tidak ada perubahan, dan juga anggota *cluster* yang dihasilkan sudah tidak berpindah-pindah.
11. Selanjutnya dikarenakan hasil nilai jarak dari setiap data ke *centroid* sudah tidak mengalami perubahan, atau jika hasil nilainya sudah sama dengan iterasi

sebelumnya, itu berarti proses *k-means* selesai pada iterasi ke- 8. Berikut pada Tabel 3.5 merupakan hasil anggota-anggota tiap *cluster*, diberikan contoh sebanyak 20 data :

Tabel 3. 4 Hasil *Cluster* Iterasi Kedelapan

| Data ke- i | C1 | C2 | C3 | Terdekat | Cluster |
|------------|----------|----------|----------|-----------|---------|
| 1 | 10.5066 | 15.12462 | 18.96596 | 10.506599 | C1 |
| 2 | 8.135817 | 18.09852 | 18.09271 | 8.1358168 | C1 |
| 3 | 15.23621 | 23.68265 | 12.46467 | 12.464675 | C3 |
| 4 | 10.5066 | 15.12462 | 18.96596 | 10.506599 | C1 |
| 5 | 13.72336 | 22.03507 | 13.76846 | 13.723361 | C1 |
| 6 | 32.63665 | 39.35672 | 17.85361 | 17.853607 | C3 |
| 7 | 6.662478 | 16.65086 | 20.36781 | 6.6624782 | C1 |
| 8 | 6.662478 | 16.65086 | 20.36781 | 6.6624782 | C1 |
| 9 | 8.720137 | 18.86763 | 19.6914 | 8.720137 | C1 |
| 10 | 11.71523 | 16.69969 | 29.50116 | 11.715229 | C1 |
| 11 | 23.07995 | 28.07218 | 10.32406 | 10.324059 | C3 |
| 12 | 30.45515 | 37.12166 | 15.82312 | 15.823123 | C3 |
| 13 | 7.919902 | 15.93116 | 24.83881 | 7.919902 | C1 |
| 14 | 8.720137 | 18.86763 | 19.6914 | 8.720137 | C1 |
| 15 | 4.792831 | 15.29959 | 20.64559 | 4.7928305 | C1 |
| 16 | 9.570194 | 14.28347 | 18.7713 | 9.5701941 | C1 |
| 17 | 15.46432 | 19.31273 | 14.39985 | 14.399849 | C3 |
| 18 | 11.71523 | 16.69969 | 29.50116 | 11.715229 | C1 |
| 19 | 23.07995 | 28.07218 | 10.32406 | 10.324059 | C3 |
| 20 | 30.45515 | 37.12166 | 15.82312 | 15.823123 | C3 |

12. Pengujian yang dilakukan dengan menggunakan 1000 data menghasilkan nilai akurasi sebesar 57.64 %. Berikut pada Tabel 3.6 diberikan beberapa contoh data perbandingan hasil dari klasifikasi perhitungan menggunakan *Microsoft Excel* dengan nilai aktual pada dataset pasien.

Tabel 3. 5 Perbandingan Hasil dari Perhitungan *Excel* dengan Data Aktual

| Data ke-i | Cluster | Data Aktual |
|-----------|---------|-------------|
| 1 | C1 | C1 |
| 2 | C1 | C1 |
| 3 | C3 | C1 |
| 4 | C1 | C1 |
| 5 | C1 | C1 |
| 6 | C3 | C1 |
| 7 | C1 | C1 |
| 8 | C1 | C1 |
| 9 | C1 | C1 |
| 10 | C1 | C1 |
| 11 | C3 | C1 |
| 12 | C3 | C1 |
| 13 | C1 | C1 |
| 14 | C1 | C1 |
| 15 | C1 | C1 |
| 16 | C1 | C1 |
| 17 | C3 | C1 |
| 18 | C1 | C1 |
| 19 | C3 | C1 |
| 20 | C3 | C1 |
| 21 | C1 | C1 |
| 22 | C1 | C1 |
| 23 | C1 | C1 |
| 24 | C3 | C1 |

Pada pembahasan selanjutnya adalah implementasi sistem. Sistem dibuat dengan menggunakan bahasa *python* dan *google colab*. Proses klasifikasi dibagi menjadi data *training* dan data *testing*. Bagian data *training* digunakan untuk

mendukung model pelatihan, sementara bagian data *testing* digunakan untuk mendukung membangun model pengujian sistem.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Skenario Pengujian

Skenario pengujian merupakan serangkaian pengujian yang dilakukan untuk menguji sistem yang telah dibangun. Dalam penelitian ini, untuk menguji tingkat akurasi metode *k-means* yang sudah dibangun dan diuji, digunakan tabel *confusion matrix* untuk menghasilkan nilai perhitungan *accuracy*, *precision*, dan *recall*. Akan dilakukan pembagian dataset menggunakan tiga skenario yang berbeda untuk memisahkan antara data *training* dan data *testing*. Pembagian dataset ini dilakukan untuk mengevaluasi pada rasio perbandingan berapa yang menghasilkan nilai akurasi paling baik. Jumlah data yang digunakan pada ketiga skenario pembagian dataset dapat dilihat pada Tabel 4.1.

Tabel 4. 1 Pembagian Dataset

| Model | Data <i>Training</i> | Data <i>Testing</i> |
|-------|----------------------|---------------------|
| A | 90% | 10% |
| B | 80% | 20% |
| C | 70% | 30% |

Data yang digunakan adalah dataset pasien kanker paru-paru yang diperoleh dari situs www.dataworld.com yang berjumlah 1000 data. Kumpulan dataset tersebut memiliki 22 atribut yang meliputi *gender*, *air pollution*, *alcohol use*, *dust allergy*, *occupational hazards*, *'genetic risk*, *chronic lung disease*, *balanced diet*, *obesity*, *smoking*, *passive smoker*, *chest pain*, *coughing of blood*, *fatigue*, *weight loss*, *'shortness of breath*, *wheezing*, *swallowing difficulty*, *clubbing of finger nails*,

frequent cold, dry cough, snoring, dan dataset tersebut memiliki kolom level sebagai label kelas yang terdiri dari kelas *high* (riwayat pasien kanker paru berstatus tinggi), *medium* (riwayat pasien kanker paru berstatus sedang), dan *low* (riwayat pasien kanker paru berstatus rendah). Label kelas pada dataset digunakan untuk membandingkan hasil prediksi yang telah dihasilkan oleh sistem. Dari 1000 dataset terdapat 365 pasien yang berstatus kanker paru tinggi, 332 pasien yang berstatus kanker paru sedang, dan 303 pasien yang berstatus kanker paru rendah.

Berdasarkan nilai pembagi data yang terdapat dalam Tabel 4.1, jumlah data *training* dan data *testing* pada setiap model menghasilkan jumlah data yang berbeda, dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Jumlah Data Setiap Model

| Model | Perbandingan Data | Jumlah Data <i>Training</i> | Jumlah Data <i>Testing</i> |
|--------------|--------------------------|--|---------------------------------------|
| A | 90:10 | 900 | 100 |
| B | 80:20 | 800 | 200 |
| C | 70:30 | 700 | 300 |

Disiapkan sebanyak 1000 data pasien, selanjutnya data tersebut akan dibagi menjadi data *training* dan data *testing* dengan perbandingan 90:10, 80:20, dan 70:30. Pengujian model A dengan perbandingan 90:10 maka didapatkan 900 data *training* dan 100 data *testing*. Pengujian model B dengan perbandingan 80:20 maka didapatkan 800 data *training* dan 200 data *testing*. Pengujian model C dengan perbandingan 70:30 maka didapatkan 700 data *training* dan 300 data *testing*. Jika sudah mendapatkan hasilnya setelah itu dihitung nilai *accuracy*, *precision*, dan *recall*.

Confusion matrix memiliki konsep yang dapat diterapkan untuk menghasilkan perhitungan *accuracy*, *precision*, dan *recall* karena *confusion matrix* memvisualisasikan hasil dari kelas aktual dan kelas prediksi dalam bentuk tabel. Berikut rumus yang digunakan untuk menghitung nilai *accuracy*, *precision*, dan *recall* (Heydarian & Doyle, 2022).

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100\% \quad (4.1)$$

$$Precision = \frac{TP}{(TP+FP)} \times 100\% \quad (4.2)$$

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \quad (4.3)$$

Keterangan :

1. (TP) *True Positive* : data pasien dengan hasil prediksi kelas i , pada data aktual juga masuk kelas i
2. (TN) *True Negative* : data pasien dengan hasil prediksi bukan kelas i , pada data aktual juga tidak masuk kelas i
3. (FP) *False Positive* : data pasien dengan hasil prediksi kelas i , pada data aktual bukan masuk kelas i
4. (FN) *False Negative* : data pasien dengan hasil prediksi bukan kelas i , pada data aktual masuk kelas i

4.2 Hasil Uji Coba

Pada sub bab ini, akan dijelaskan hasil pengujian sistem berdasarkan uji coba yang dilakukan untuk mengevaluasi akurasi penggunaan metode *k-means*. Nilai akurasi yang dihasilkan dalam uji coba model didasarkan pada perbandingan antara nilai y_{test} (kelas aktual) dan y_{pred} (kelas yang diprediksi). Langkah selanjutnya adalah melakukan evaluasi terhadap sistem yang telah dibangun untuk menilai performa metode *k-means* dalam mengklasifikasikan data pasien kanker paru, khususnya dalam hal akurasi, yaitu digunakan *confusion matrix* yang

menggambarkan hasil klasifikasi dengan membandingkan data aktual dan data prediksi dari proses klasifikasi yang telah dibangun.

4.2.1 Hasil Pengujian Model A

Disiapkan sebanyak 1000 data pasien, kemudian data tersebut dipisahkan menjadi data *training* dan data *testing* dengan perbandingan rasio 90:10. Digunakan sebanyak 900 data *training* dan 100 data *testing*. Berikut pada Tabel 4.3 adalah tabel *confusion matrix* yang menunjukkan hasil klasifikasi dari pengujian model A.

Tabel 4. 3 Counfusion Matrix Pengujian Model A

| Data Aktual | Hasil Prediksi | | |
|-------------|----------------|-----|--------|
| | High | Low | Medium |
| High | 34 | 0 | 4 |
| Low | 2 | 26 | 0 |
| Medium | 6 | 15 | 13 |

Untuk perhitungan *accuracy*, *precision*, dan *recall* dari hasil pengujian model A adalah sebagai berikut :

$$Accuracy = \frac{34+26+13}{100} \times 100\% = 73\%$$

Kemudian pada Tabel 4.4 ditunjukkan nilai *precision* dan *recall* dari setiap kelas, dan ditunjukkan juga nilai rata-ratanya.

Tabel 4. 4 Nilai *Recall* dan *Precision* Setiap Kelas untuk Pengujian Model A

| Kelas | <i>Precision</i> | <i>Recall</i> |
|--------------|------------------|---------------|
| High | 80.95 % | 89.47% |
| Low | 63.41% | 92.85% |
| Medium | 76.47% | 38.23% |
| Total | 73.61% | 73.51% |

Hasil dari pengujian model A menghasilkan *accuracy* 73%, untuk kelas *high* menghasilkan *precision* 80.95% dan *recall* 89.47%, untuk kelas *low* menghasilkan *precision* 63.41% dan *recall* 92.85%, dan untuk kelas *medium* menghasilkan *precision* 76.47% dan *recall* 38.23%.

4.2.2 Hasil Pengujian Model B

Disiapkan sebanyak 1000 data pasien, kemudian data tersebut dipisahkan menjadi data *training* dan data *testing* dengan perbandingan rasio 80:20. Digunakan sebanyak 800 data *training* dan 200 data *testing*. Berikut pada Tabel 4.5 adalah tabel *confusion matrix* yang menunjukkan hasil klasifikasi dari pengujian model B.

Tabel 4. 5 Counfusion Matrix Pengujian Model B

| Data Aktual | Hasil Prediksi | | |
|-------------|----------------|-----|--------|
| | High | Low | Medium |
| High | 60 | 0 | 20 |
| Low | 4 | 46 | 0 |
| Medium | 14 | 35 | 21 |

Untuk perhitungan *accuracy*, *precision*, dan *recall* dari hasil pengujian model B adalah sebagai berikut :

$$Accuracy = \frac{60+46+21}{200} \times 100\% = 63.5\%$$

Kemudian pada Tabel 4.6 ditunjukkan nilai *precision* dan *recall* dari setiap kelas, dan ditunjukkan juga nilai rata-ratanya.

Tabel 4. 6 Nilai Recall dan Precision Setiap Kelas untuk Pengujian Model B

| Kelas | <i>Precision</i> | <i>Recall</i> |
|-------|------------------|---------------|
| High | 76.92 % | 75% |
| Low | 56.79% | 92% |

| | | |
|--------------|--------|--------|
| Medium | 51.21% | 30% |
| Total | 61.64% | 65.66% |

Hasil dari pengujian model B menghasilkan *accuracy* 63.5%, untuk kelas *high* menghasilkan *precision* 76.92% dan *recall* 75%, untuk kelas *low* menghasilkan *precision* 56.79% dan *recall* 92%, dan untuk kelas *medium* menghasilkan *precision* 51.21% dan *recall* 30%.

4.2.3 Hasil Pengujian Model C

Disiapkan sebanyak 1000 data pasien, kemudian data tersebut dipisahkan menjadi data *training* dan data *testing* dengan perbandingan rasio 70:30. Digunakan sebanyak 700 data *training* dan 300 data *testing*. Berikut pada Tabel 4.7 adalah tabel *confusion matrix* yang menunjukkan hasil klasifikasi dari pengujian model C.

Tabel 4. 7 Counfusion Matrix Pengujian Model C

| Data Aktual | Hasil Prediksi | | |
|-------------|----------------|-----|--------|
| | High | Low | Medium |
| High | 93 | 0 | 26 |
| Low | 6 | 78 | 0 |
| Medium | 18 | 46 | 33 |

Untuk perhitungan *accuracy*, *precision*, dan *recall* dari hasil pengujian model C adalah sebagai berikut :

$$Accuracy = \frac{93+78+33}{300} \times 100\% = 68\%$$

Kemudian pada Tabel 4.8 ditunjukkan nilai *precision* dan *recall* dari setiap kelas, dan ditunjukkan juga nilai rata-ratanya.

Tabel 4. 8 Nilai Recall dan Precision Setiap Kelas untuk Pengujian Model C

| Kelas | <i>Precision</i> | <i>Recall</i> |
|--------------|-------------------------|----------------------|
| High | 79.48% | 78.15% |
| Low | 62.90% | 92.85% |
| Medium | 55.93% | 34.02% |
| Total | 66.10% | 68.34% |

Hasil dari pengujian model C menghasilkan *accuracy* 68%, untuk kelas *high* menghasilkan *precision* 79.48% dan *recall* 78.15%, untuk kelas *low* menghasilkan *precision* 62.90% dan *recall* 92.85%, dan untuk kelas *medium* menghasilkan *precision* 55.93% dan *recall* 34.02%.

4.2.4 Pengujian *K-Fold Cross Validation*

Digunakan juga pengujian *k-fold cross validation* untuk evaluasi metode *k-means* dalam melakukan klasifikasi pada dataset pasien kanker paru. Pada pengujian metode *k-fold cross validation*, data dibagi menjadi *k* subset atau *fold* yang sama ukurannya. Proses pengujian dilakukan sebanyak *k* kali, dimana setiap *fold* digunakan sebagai data *testing* satu kali dan *fold* yang lain sebagai data *training*. Akurasi model dihitung sebagai rata-rata dari akurasi yang diperoleh pada setiap iterasi *k-fold*. Pada penelitian ini nilai *k* yang digunakan adalah 5, dalam Tabel 4.9 disajikan perbandingan hasil akurasi metode *k-means* berdasarkan nilai parameter *k* dari 1 sampai 5.

Tabel 4. 9 Hasil Akurasi Menggunakan Pengujian *K-Fold Cross Validation*

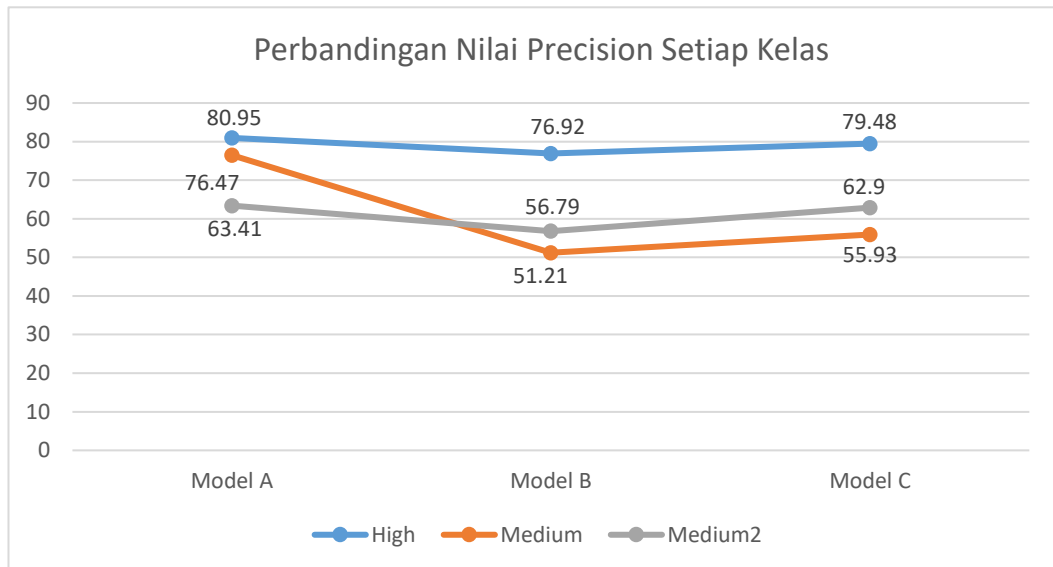
| Parameter- k | Akurasi |
|----------------|---------|
| 1 | 45% |
| 2 | 50.5% |
| 3 | 53% |
| 4 | 39.5% |
| 5 | 32% |

Berdasarkan hasil yang diperoleh yang disajikan pada Tabel 4.9 dapat diketahui bahwa pada nilai parameter $k=1$ diperoleh nilai akurasi sebesar 45%, parameter $k=2$ diperoleh nilai akurasi sebesar 50.5%, parameter $k=3$ diperoleh nilai akurasi sebesar 53%, parameter $k=4$ diperoleh nilai akurasi sebesar 39.5%, dan pada parameter $k=5$ diperoleh nilai akurasi sebesar 32%. Maka yang memiliki nilai akurasi paling tinggi dihasilkan oleh parameter $k=3$ yaitu sebesar 53%.

4.3 Pembahasan

Berdasarkan hasil uji coba yang dilakukan terhadap dataset pasien kanker paru yang terdiri dari 1000 sampel dan 22 atribut, dengan kolom level sebagai label kelas yang memiliki kelas *high*, *medium*, dan *low*. Dari penggunaan data yang digunakan dan uji coba yang dilakukan, dapat diketahui bahwa nilai akurasi tertinggi didapatkan melalui pembagian data *training* dan data *testing* pada model A dengan rasio perbandingan 90:10, nilai akurasi yang dihasilkan yaitu sebesar 73%. Selanjutnya dari hasil yang diperoleh pada semua model menunjukkan bahwa model cenderung dapat memprediksi kelas atau kategori *high*, dapat dilihat dari

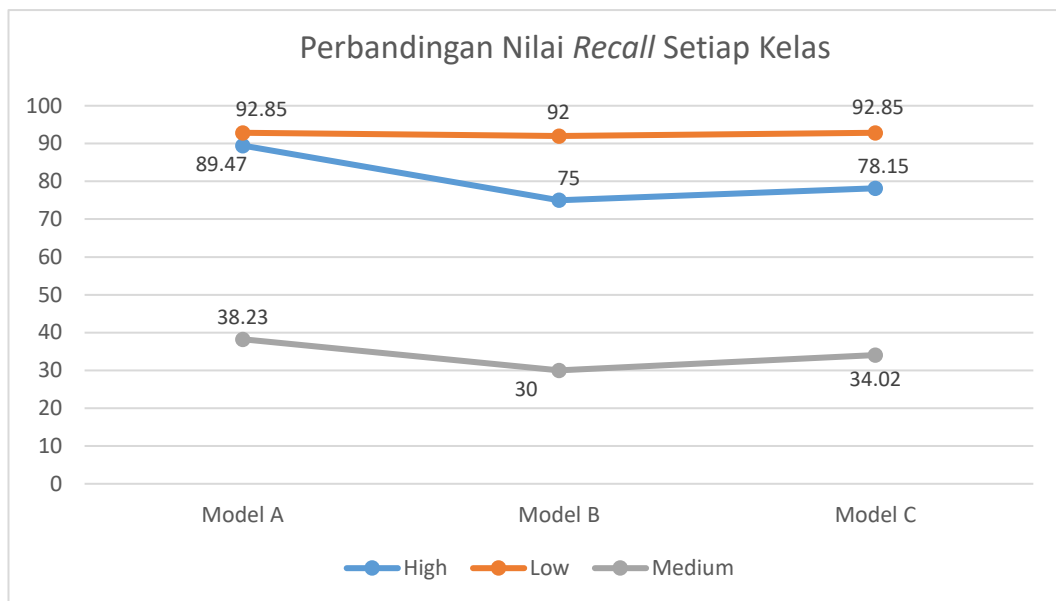
perbandingan hasil nilai *precision* pada setiap model. Berikut pada Gambar 4.1 merupakan hasil perbandingan nilai *precision* yang disajikan dalam bentuk grafik.



Gambar 4. 1 Grafik Perbandingan Nilai *Precision* Setiap Kelas

Hasil perbandingan nilai *precision* dari ketiga model tersebut menunjukkan bahwa model cenderung dapat memprediksi kelas *high*, hal ini terlihat dari nilai *precision* pada kelas *high* di setiap hasil uji coba memiliki nilai tertinggi dibandingkan kelas lainnya. Dapat diketahui pada pengujian model A dengan rasio pembagian data *training* dan data *testing* 90:10 menghasilkan nilai *precision* yang paling baik, yaitu sebesar 80.95%.

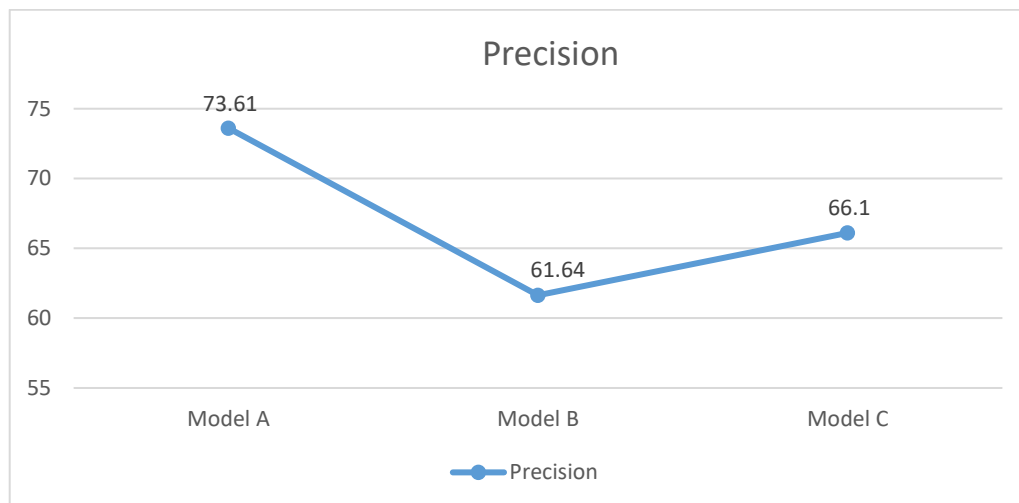
Selanjutnya pembahasan mengenai hasil nilai *recall* yang diperoleh. *Recall* digunakan untuk mengukur sejauh mana model dapat mengidentifikasi data positif dengan benar, untuk *recall* fokus pada jumlah data positif yang berhasil ditemukan dari semua data positif yang ada. Berikut dapat dilihat pada Gambar 4.2 merupakan perbandingan hasil nilai *recall* dari ketiga model.



Gambar 4. 2 Grafik Perbandingan Nilai *Recall* Setiap Kelas

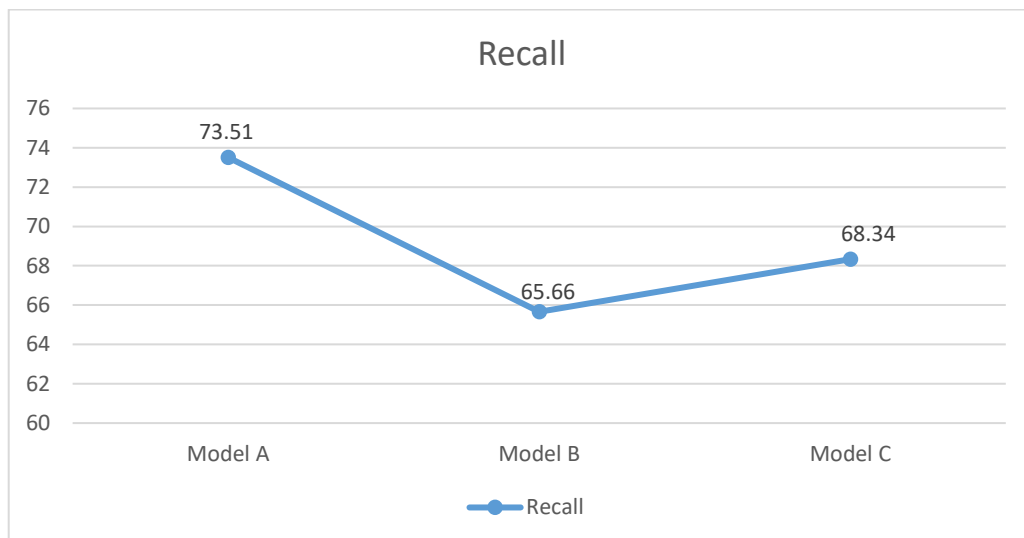
Berdasarkan Gambar 4.2 dapat dilihat grafik nilai *recall* setiap kelas pada setiap model. Hasil perbandingan nilai *recall* dari uji coba yang dilakukan pada ketiga model tersebut menunjukkan bahwa pada pengujian model A dan model C dengan rasio pembagian data *training* dan data *testing* 90:10 dan 70:30 menghasilkan nilai *recall* yang paling baik, yaitu sebesar 92.85%. Sedangkan dari semua model ternyata nilai *recall* yang paling buruk adalah pada kelas *medium*, yaitu pada model A nilai *recall* yang dihasilkan sebesar 38.23%, pada model B nilai *recall* yang dihasilkan sebesar 30%, dan pada model C nilai *recall* yang dihasilkan

sebesar 34.02%. Selanjutnya pembahasan mengenai perbandingan hasil rata-rata nilai *precision* dan *recall* dari hasil uji coba pada ketiga model.



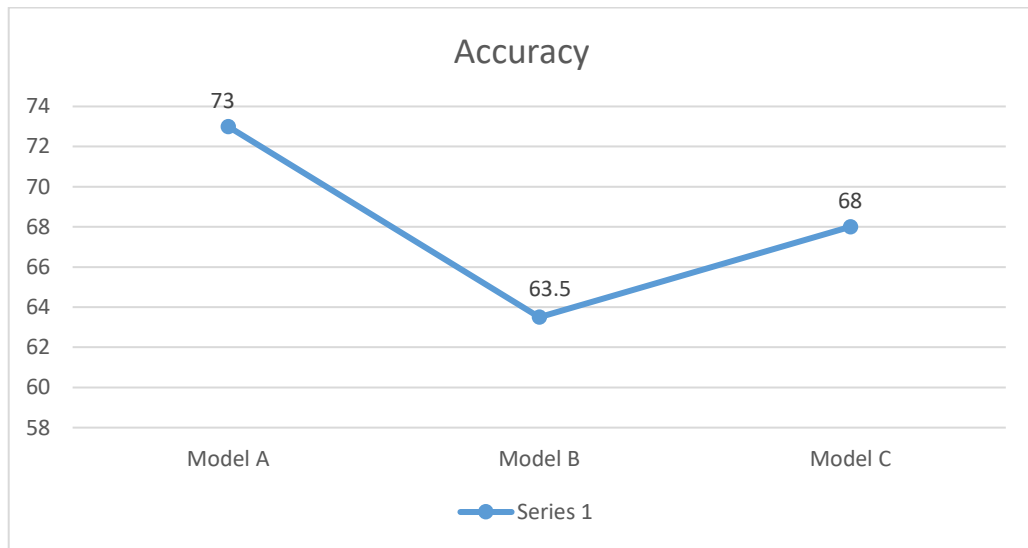
Gambar 4. 3 Perbandingan Nilai Rata-rata *Precision* dari Ketiga Model

Hasil akhir nilai *precision* dari uji coba pada ketiga model dapat dilihat pada Gambar 4.3 yang disajikan dalam bentuk grafik, menunjukkan bahwa hasil rata-rata nilai *precision* dari model A dengan rasio perbandingan data *training* dan data *testing* 90:10 menghasilkan nilai *precision* yang paling baik, yaitu sebesar 73.61%. *Precision* digunakan untuk mengukur sejauh mana model dapat mengklasifikasikan data sebagai positif secara akurat. Semakin tinggi nilai *precision* yang diperoleh, maka semakin sedikit data negatif yang salah diklasifikasikan sebagai positif. Sebagai contoh hasil klasifikasi dari suatu data diidentifikasi termasuk kelas *high*, dan benar pada data aktualnya juga termasuk kelas *high*.



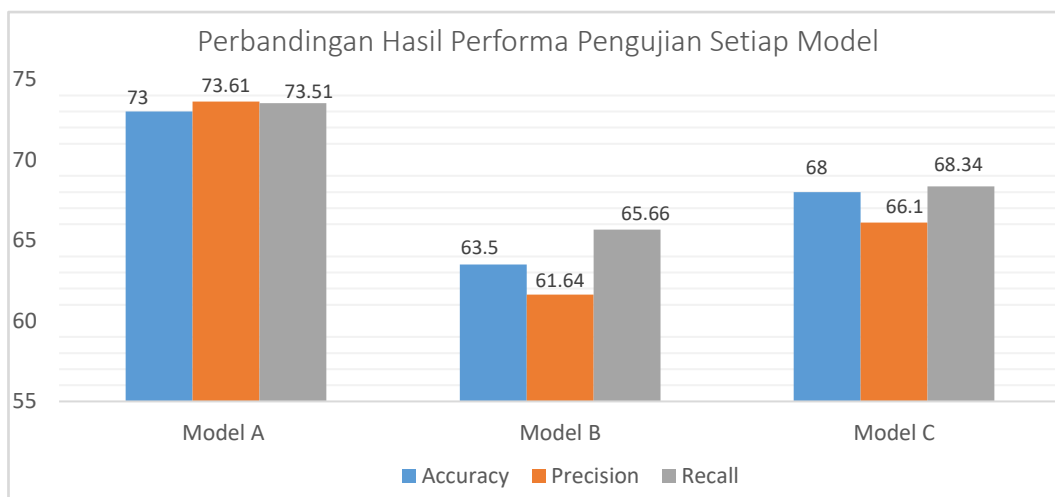
Gambar 4. 4 Perbandingan Nilai Rata-rata *Recall* dari Ketiga Model

Hasil akhir rata-rata nilai *recall* dari uji coba yang dilakukan pada ketiga model dapat dilihat pada Gambar 4.4 yang disajikan dalam bentuk garafik, menunjukkan bahwa hasil rata-rata nilai *precision* dari model A dengan rasio perbandingan data *training* dan data *testing* 90:10 menghasilkan nilai *recall* yang paling baik, yaitu sebesar 73.61%. Pada konteks klasifikasi penyakit, *recall* memberikan informasi seberapa baik model dapat mengklasifikasi dan mengidentifikasi pasien yang memang termasuk ke dalam label kelas yang sebenarnya. Jika nilai *recall* yang dihasilkan rendah, berarti model menghasilkan lebih banyak *false negative*, sebagai contoh seharusnya pada data aktual pasien sebenarnya termasuk kelas *high* tapi dianggap ke dalam kelas *low* oleh model. Selanjutnya pembahasan mengenai perbandingan hasil nilai *accuracy* yang sudah dilakukan uji coba pada ketiga model.



Gambar 4. 5 Perbandingan Nilai Akurasi setiap Model

Hasil uji coba dari model A menunjukkan nilai akurasi yang didapat sebesar 73%. Hasil uji coba dari model B menunjukkan nilai akurasi yang didapat sebesar 63.5%. Hasil uji coba dari model C menunjukkan nilai akurasi yang didapat sebesar 68%. Selanjutnya, akan dibahas perbandingan hasil pengujian akhir dari ketiga model. Perbandingan hasil tersebut dapat dilihat pada Gambar 4.6.



Gambar 4. 6 Diagram Perbandingan Performa Setiap Model

Hasil pengujian pada model A menunjukkan bahwa nilai akurasi yang diperoleh adalah 73%, *precision* sebesar 73.61%, dan *recall* sebesar 73.51%. Sementara itu, hasil pengujian pada model B menghasilkan nilai akurasi sebesar 63.5%, *precision* sebesar 61.64%, dan *recall* sebesar 65.66%. Selanjutnya, hasil pengujian pada model C menunjukkan nilai akurasi sebesar 68%, *precision* sebesar 66.1%, dan *recall* sebesar 68.34%. Berdasarkan perbandingan hasil performa pengujian setiap model pada Gambar 4.6 didapatkan bahwa uji coba yang dilakukan pada model A dengan menggunakan 1000 data dan rasio perbandingan data *training* dan data *testing* 90:10 menghasilkan nilai akurasi yang paling baik, yaitu dengan nilai *accuracy* sebesar 73%, *precision* 73.61%, dan *recall* 74.51%.

Peneliti juga telah melakukan pengujian dengan mengurangi jumlah data yang digunakan, jumlah data yang digunakan yaitu sebanyak 800 sampel. Pengurangan dataset ini dilakukan untuk melihat perbandingan nilai akurasi yang diperoleh, apakah dengan mengurangi jumlah data akan menghasilkan nilai akurasi yang lebih baik jika dibandingkan dengan hasil klasifikasi menggunakan 1000 sampel. Pada pengujian 800 sampel juga dilakukan uji coba dengan ketiga model, dari hasil uji coba dengan menggunakan ketiga model tersebut diperoleh nilai akurasi yang paling baik yaitu pada model B dengan rasio perbandingan data *training* dan data *testing* 80:20. Pada model B dengan menggunakan 800 sampel menunjukkan nilai akurasi yang diperoleh sebesar 53.57%, nilai *precision* sebesar 55.54%, dan nilai *recall* sebesar 54.34%.

Terdapat beberapa kelompok klasifikasi berdasarkan rentang nilai, rentang nilai 90% - 100% digolongkan sebagai klasifikasi sangat baik, rentang nilai 80% -

90% digolongkan sebagai klasifikasi baik, rentang nilai 70% - 80% digolongkan sebagai klasifikasi cukup, rentang nilai 60% - 70% digolongkan sebagai klasifikasi kurang baik, dan rentang nilai 50% - 60% digolongkan sebagai klasifikasi gagal. (Gorunescu, 2011). Berdasarkan tingkatan nilai klasifikasi tersebut, hasil uji coba dengan 1000 sampel pada rasio perbandingan data *training* dan data *testing* 90:10 (model A) yang memiliki nilai akurasi sebesar 73% diklasifikasikan cukup, pada model B memiliki nilai akurasi sebesar 63.5% maka diklasifikasikan kurang baik, dan pada model C memiliki nilai akurasi sebesar 68% maka diklasifikasikan kurang baik juga. Sedangkan nilai akurasi dengan menggunakan *k-fold cross validation* paling tinggi pada parameter $k=3$ yaitu sebesar 53%. Berdasarkan hasil nilai akurasi yang diperoleh, terdapat beberapa faktor yang bisa mempengaruhi perolehan nilai akurasi.

Pada hasil pengujian menggunakan *k-fold cross validation* diperoleh nilai-nilai akurasi yang tergolong gagal, hal tersebut bisa terjadi dikarenakan metode *k-means* sangat bergantung pada inisialisasi titik *centroid*. Jika titik *centroid* awal yang dipilih secara random tidak mewakili dengan baik pola data yang sebenarnya, maka hasil klasifikasi dapat menjadi kurang akurat.

Hasil yang diperoleh dengan membandingkan perbedaan jumlah dataset yang digunakan untuk dilakukan uji coba ternyata tidak menunjukkan perbaikan yang signifikan terhadap hasil akurasi yang diperoleh dari ketiga model menggunakan metode *k-means*. Hal ini dikarenakan *k-means* merupakan metode pengelompokan *unsupervised learning*, yang mana dalam pengelompokannya tidak memerlukan informasi label dari data. Tetapi dalam tugas klasifikasi kita memiliki

label yang diberikan dan ingin melatih model untuk mengidentifikasi kelas tertentu, oleh karena itu pada metode *k-means* tidak mempertimbangkan informasi label ini secara langsung, sehingga tidak menghasilkan model yang sesuai untuk tugas klasifikasi.

Faktor lainnya yang mempengaruhi performa pada klasifikasi menggunakan metode *k-means* adalah data yang terlalu kompleks. Ketika proses klasifikasi, jumlah variabel pada data sangat mempengaruhi hasil prediksi. Karena metode *k-means* hanya mempertimbangkan jarak antar titik, hal ini yang dapat menghasilkan klasifikasi yang salah jika terdapat banyak variabel dalam data.

Tujuan dari penelitian ini adalah untuk mengklasifikasikan penyakit kanker paru pada pasien yang memiliki faktor risiko, sehingga dengan adanya sistem ini dapat membantu pasien kanker paru untuk mengetahui status penyakitnya apakah tergolong rendah, sedang atau tinggi. Walaupun dari sistem yang sudah dibangun perlu adanya pengembangan atau perbaikan lagi supaya menghasilkan performa sistem yang lebih baik, sehingga sistem nantinya dapat menjadi lebih bermanfaat, seperti dapat mengurangi risiko kesalahan dalam pengambilan keputusan pada penanganan pasien yang menderita kanker paru.

Tidak semua orang selalu merasakan kesehatan tubuh yang optimal, terkadang beberapa orang menghadapi beberapa penyakit, baik yang menyerang organ luar maupun organ dalam. Terkait dengan hal tersebut, Allah SWT tidak menginginkan hamba-Nya pasrah ketika sakit, namun Allah SWT meminta dan mewajibkan kita untuk berusaha sebaik mungkin agar sembuh dari penyakit yang kita alami. Seperti yang telah disampaikan oleh Nabi Muhammad SAW kepada

para sahabat dan umatnya, disarankan untuk mencari pengobatan ketika mengalami sakit, karena setiap penyakit dapat ditemukan obatnya, seperti dalam hadist Rasulullah SAW :

حَدَّثَنَا هَارُونُ بْنُ مَعْرُوفٍ وَأَبُو الطَّاهِرِ وَأَحْمَدُ بْنُ عِيسَى قَالُوا حَدَّثَنَا ابْنُ وَهْبٍ أَخْبَرَنِي عَمْرُو وَهُوَ ابْنُ الْحَارِثِ عَنْ عَبْدِ رَبِّهِ بْنِ سَعِيدٍ عَنْ أَبِي الزُّبَيْرِ عَنْ جَابِرٍ عَنْ رَسُولِ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ أَنَّهُ قَالَ لِكُلِّ دَاءٍ دَوَاءٌ فَإِذَا أُصِيبَ دَوَاءُ الدَّاءِ بَرَأَ بِإِذْنِ اللَّهِ عَزَّ وَجَلَّ

“Telah menceritakan kepada kami Harun bin Ma’ruf dan Abu Ath Thahir serta Ahmad bin ‘Isa mereka berkata; Telah menceritakan kepada kami Ibnu Wahb; Telah mengabarkan kepadaku ‘Amru, yaitu Ibnu al-Harits dari ‘Abdu Rabbih bin Sa’id dari Abu Az Zubair dari Jabir dari Rasulullah shallallahu ‘alaihi wasallam, beliau bersabda: “Setiap penyakit ada obatnya. Apabila ditemukan obat yang tepat untuk suatu penyakit, akan sembuhlah penyakit itu dengan izin Allah ‘azza wajalla.” (HR Muslim).

Ibnu Qayyim dalam sebuah kitabnya menyampaikan bahwa pada kalimat “Setiap penyakit ada obatnya” bersifat secara umum. Allah menciptakan obat-obatan untuk menyembuhkan berbagai penyakit. Hadits tersebut mengisyaratkan perintah untuk berobat. Menurut Ibnu Qayyim, berobat sama sekali tidak bertentangan dengan tawakkal.

Terkait dengan tujuan sistem ini, yaitu melakukan klasifikasi atau pengelompokan berdasarkan perhitungan dan aturan yang ditentukan, dalam Islam segala sesuatu yang Allah ciptakan juga memiliki aturan yang pasti dan ukuran yang ditetapkan. Hal ini sejalan dengan firman Allah *subhanahu wa ta’ala* dalam surah Al-Qamar ayat 49 :

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ

“Sesungguhnya Kami menciptakan segala sesuatu dengan qadar (ukuran,

aturan)” (Q.S. Al-Qamar : 49)

Dalam Al-Qur’an mengajarkan tentang pentingnya mengenal dan memahami sifat-sifat ciptaan Allah SWT serta memperhatikan tujuan di balik penciptaan tersebut. Dalam konteks pengklasifikasian penyakit, ini dapat berarti mengelompokkan entitas berdasarkan sifat atau karakteristik tertentu yang mempengaruhi tingkat keparahan penyakit yang diderita. Pada penelitian ini mengelompokkan pasien kanker paru berdasarkan faktor risiko, contohnya apabila beberapa faktor risiko yang dimilikinya tergolong pada level tinggi maka ia akan dikelompokkan ke dalam level *high* (riwayat pasien kanker paru berstatus tinggi).

Menurut tafsir Ibnu Katsir, ayat 49 dari surah Al-Qamar yang telah dijelaskan sebelumnya juga memiliki makna yang sejalan dengan apa yang Allah *subhanahu wa ta’ala* sampaikan dalam ayat 2 dari surah Al-Furqan :

وَحَلَقَ كُلَّ شَيْءٍ فَقَدَرَهُ تَقْدِيرًا

“Dan Dia menciptakan segala sesuatu, dan Dia menetapkan ukuran-ukurannya dengan serapi-rapinya.” (Q.S. Al-Furqan : 2)

Dalam konteks klasifikasi, ayat tersebut dapat dicerminkan dengan pentingnya mengelompokkan entitas berdasarkan kriteria yang jelas dan teratur.

Terdapat juga firman Allah *subhanahu wa ta’ala* dalam surah Al-A’la ayat 1-3 :

سَبِّحْ اسْمَ رَبِّكَ الْأَعْلَى . الَّذِي خَلَقَ فَسَوَّى . وَالَّذِي قَدَّرَ فَهَدَى

“Sucikanlah nama Tuhanmu Yang Maha Tinggi, yang menciptakan dan yang menyempurnakan (penciptaan-Nya) dan yang menentukan kadar (masing-masing) dan memberi petunjuk.” (Q.S. Al-A’la: 1-3)

Berdasarkan ketiga ayat diatas, diketahui bahwa Allah *subhanahu wa ta'ala* telah menentukan ukuran bagi setiap makhluk-Nya dan memberikan petunjuk kepada makhluk-Nya yang lain. Dengan mengintegrasikan prinsip-prinsip Islam seperti ayat-ayat tersebut, bahwa semua ada ukurannya, dalam pelaksanaan klasifikasi kita dapat memastikan bahwa penelitian ini mengikuti prinsip-prinsip yang dengan sesuai dengan nilai-nilai Islam.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah pengujian terhadap pengklasifikasian dataset pasien kanker paru menggunakan metode *k-means* telah dilakukan maka dapat disimpulkan bahwa pengujian menggunakan 1000 sampel pada uji coba model A dengan rasio perbandingan data *training* dan data *testing* yaitu 90:10 menghasilkan nilai akurasi yang paling baik, yaitu didapatkan nilai *accuracy* sebesar 73%, *precision* sebesar 73.61%, dan *recall* sebesar 73.51%. Sedangkan nilai akurasi dengan menggunakan *k-fold cross validation* paling tinggi pada parameter $k=3$ yaitu sebesar 53%.

Faktor yang mempengaruhi performa pada klasifikasi penyakit kanker paru menggunakan metode *k-means* adalah data yang terlalu kompleks. Metode *k-means* cenderung menghasilkan klasifikasi yang kurang baik pada data yang sangat kompleks, terutama apabila terdapat banyak variabel atau atribut dalam data. Karena metode *k-means* hanya mempertimbangkan jarak antar titik, hal ini yang dapat menghasilkan klasifikasi yang salah jika terdapat banyak variabel dalam data. Selain itu, pada hasil pengujian menggunakan *k-fold cross validation* menunjukkan nilai-nilai akurasi yang tergolong gagal, hal tersebut bisa terjadi dikarenakan metode *k-means* sangat bergantung pada inisialisasi titik *centroid*. Jika titik *centroid* awal yang dipilih secara random tidak mewakili dengan baik pola data yang sebenarnya, maka hasil klasifikasi dapat menjadi kurang akurat.

5.2 Saran

Berdasarkan hasil pengujian terhadap pengklasifikasian data pasien kanker paru menggunakan metode *K-Means*, penulis mengakui adanya kekurangan dan kelemahan dalam penelitian ini dan menyadari perlunya pengembangan lebih lanjut guna mencapai hasil yang lebih akurat. Berikut adalah beberapa saran dari penulis untuk penelitian selanjutnya :

1. Supaya menghasilkan nilai akurasi yang lebih optimal maka dapat dilakukan uji coba dengan mereduksi variabel.
2. Akan lebih baik jika metode yang digunakan merupakan metode *supervised learning*, yang mana metode tersebut biasa digunakan dalam mengklasifikasikan data atau memprediksi hasil secara akurat. Beberapa contoh metode supervised learning yang dapat digunakan yaitu metode *K-Nearest Neighbor*, *Support Vector Machine*, *Naive Bayes*, dan *Neural Networks*.

DAFTAR PUSTAKA

- Abrianto, P. M. C. (2018). Penerapan Metode K-Means Clustering Untuk Pengelompokan Pasien Penyakit Liver. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2(2), 247–255.
- Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(1), 186–195. <https://doi.org/10.30812/matrik.v19i1.529>
- Aliyah, N., Pranggono, E., & Andriyoko, B. (2016). Kanker Paru: Sebuah Kajian Singkat. *Indonesian Journal Chest And Emergency Medicine*, 4(1), 28–32.
- Amalita, N., & Dewi, M. P. (2021). Faktor-Faktor Risiko yang Mempengaruhi Kanker Paru-Paru dengan Menggunakan Analisis Regresi Logistik. *UNPjoMath*, 4(1), 38–42.
- Arifin, T., & Syalwah, S. (2020). Prediksi Keberhasilan Immunotherapy Pada Penyakit Kutil Dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Responsif: Riset Sains Dan Informatika*, 2(1), 38–43. <https://doi.org/10.51977/jti.v2i1.177>
- Atthalla, I. N., Jovandy, A., & Habibie, H. (2018). Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor. *Prosiding Annual Research Seminar*, 4(1), 148–151.
- Barta, J. A., Powell, C. A., & Wisnivesky, J. P. (2019). Global epidemiology of lung cancer. *Annals of Global Health*, 85(1), 1–16. <https://doi.org/10.5334/aogh.2419>
- Bashit, N., Prasetyo, Y., & Suprayogi, A. (2019). Klasifikasi Berbasis Objek untuk Pemetaan Penggunaan Lahan menggunakan Citra SPOT 5 di Kecamatan Ngaglik. *Teknik*, 40(2), 122. <https://doi.org/10.14710/teknik.v39i3.23050>
- Buana, I., & Harahap, D. A. (2022). Asbestos, Radon Dan Polusi Udara Sebagai Faktor Resiko Kanker Paru Pada Perempuan Bukan Perokok. *AVERROUS: Jurnal Kedokteran Dan Kesehatan Malikussaleh*, 8(1), 1. <https://doi.org/10.29103/averrous.v8i1.7088>
- Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, 5(1), 10. <https://doi.org/10.19184/isj.v5i1.17071>
- Febriani, A., & Furqon, A. (2020). Metastasis Kanker Paru. *Jurnal Respirasi*, 4(3), 94. <https://doi.org/10.20473/jr.v4-i.3.2018.94-101>
- Gorunescu, F. (2011). *Data Mining (Concepts, Models and Techniques)*. Springer

Singapore.

- Heydarian, M., & Doyle, T. E. (2022). *MLCM: Multi-Label Confusion Matrix*. 19083–19095.
- Irma Devi Lestari. (2016). 196927-ID-klasifikasi-online-dan-google. *Iqra'*, 83–94.
- Li, X. X., Li, B., Tian, L. F., & Zhang, L. (2018). Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. *IET Image Processing*, 12(7), 1253–1264. <https://doi.org/10.1049/iet-ipr.2016.1014>
- Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110–118.
- Pangribowo, S. (2019). Beban Kanker di Indonesia. *Pusat Data Dan Informasi Kesehatan Kementerian Kesehatan RI*, 1–16.
- Prabowo, Y., Studi, P., Informatika, T., Informasi, F. T., Luhur, U. B., Studi, P., Komputer, S., Informasi, F. T., & Luhur, U. B. (2022). *Diagnosis Kanker Paru-Paru dengan Sistem Fuzzy*. 10(1), 19–28. <https://doi.org/10.32832/kreatif.v10i1.6317>
- Prasetio, R. T., & Susanti, S. (2019). Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor. *Jurnal Responsif*, 1(1), 64–69. <http://ejurnal.univbsi.id/index.php/jti>
- Pribadi, W. W., Yunus, A., & Sartika Wiguna, A. (2022). Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang. *Jurnal Mahasiswa Teknik Informatika*, 6(2), 493–500.
- Purnamaningsih, C., Saptono, R., & Aziz, A. (2016). Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA. *Jurnal Teknologi & Informasi ITSmart*, 3(1), 27. <https://doi.org/10.20961/its.v3i1.644>
- Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 15–21. <https://doi.org/10.47970/siskom-kb.v4i1.169>
- Risnawati, R., Pradjoko, I., & Wati, F. F. (2020). Nutrisi pada Penderita Kanker Paru. *Jurnal Respirasi*, 5(3), 91. <https://doi.org/10.20473/jr.v5-i.3.2019.91-100>
- Salam, A., Adiatma, D., & Zeniarja, J. (2020). Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS. *JOINS (Journal of Information System)*, 5(1), 62–68. <https://doi.org/10.33633/joins.v5i1.3350>
- Silitonga, parasian D., & Morina, irene sri. (2017). Klusterisasi Pola Penyebaran

Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering. *Jurnal TIMES*, VI(Vol 6, No 2 (2017)), 22–25. <http://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/view/584>

Subrata, G. (2019). Klasifikasi Bahan Pustaka. *Pustakawan Perpustakaan, Ddc*, 1–13.

Toloza, E. M., Harpole, L., Detterbeck, F., & McCrory, D. C. (2003). Invasive staging of non-small cell lung cancer: A review of the current evidence. *Chest*, 123(1 SUPPL.), 157S-166S. <https://doi.org/10.1378/chest.123.1>

Yudi Agusta. (2007). K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem Dan Informatika*, 3(Februari), 47–60.

LAMPIRAN

LAMPIRAN

Lampiran I Hasil Pengujian Model

Model A (90:10)

| data_test | actual | prediction |
|---|--------|------------|
| [1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [4, 5, 6, 6, 5, 4, 6, 6, 6, 6, 6, 5, 3, 2, 4, 3, 1, 7, 5, 6] | 2 | 0 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 3, 2, 1, 3, 2, 1, 2, 7, 6, 2, 2, 2, 2, 3, 2, 1, 2, 3, 2, 3] | 1 | 1 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [6, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 3, 6, 7, 7, 7, 7, 2, 4, 3, 7, 4, 2, 1, 3, 2, 2, 1, 2, 5, 1] | 1 | 0 |
| [2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [6, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |

| data_test | actual | prediction |
|---|--------|------------|
| [6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |
| [3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [3, 2, 1, 3, 2, 4, 5, 1, 6, 2, 3, 2, 4, 3, 2, 1, 2, 4, 2, 3, 2] | 1 | 1 |
| [5, 6, 6, 5, 6, 5, 6, 5, 8, 5, 5, 5, 4, 3, 6, 2, 1, 2, 1, 6, 2] | 2 | 0 |
| [6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 6, 2, 3, 6, 6, 6, 4, 6, 8, 7, 6, 5, 5, 4, 6, 5, 4, 6, 5, 7] | 0 | 0 |
| [3, 1, 1, 1, 2, 3, 4, 1, 3, 2, 4, 3, 2, 5, 2, 1, 2, 3, 4, 5, 2] | 1 | 1 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 2, 3, 4, 4, 3, 2, 1, 3, 2, 1, 2, 1, 3, 3, 2, 1, 3, 2, 1, 1] | 1 | 1 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [5, 2, 3, 1, 2, 3, 5, 2, 2, 5, 1, 3, 1, 1, 1, 1, 1, 3, 2, 4, 2] | 1 | 1 |
| [6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |

| data_test | actual | prediction |
|---|--------|------------|
| [1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [3, 2, 4, 2, 3, 2, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 1, 2, 1, 2, 1, 2, 3, 2, 4, 2, 1, 3, 4, 5, 1, 3, 2, 1, 2, 2] | 1 | 1 |
| [1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [6, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [8, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [6, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [6, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |

Model B (80:20)

| data_test | actual | prediction |
|--|--------|------------|
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 2, 3, 4, 5, 1, 2, 3, 4] | 1 | 1 |
| [2, 4, 5, 6, 5, 6, 5, 5, 5, 6, 6, 6, 5, 1, 4, 3, 2, 4, 6, 2, 4, 1] | 2 | 0 |
| [2, 3, 2, 2, 1, 1, 1, 1, 1, 4, 2, 3, 6, 2, 1, 2, 3, 4, 2, 1, 1, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [1, 2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 2, 3, 4, 5, 1, 2, 3, 4] | 1 | 1 |
| [1, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 3, 2, 1, 3, 2, 4, 5, 1, 6, 2, 3, 2, 4, 3, 2, 1, 2, 4, 2, 3, 2] | 1 | 1 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 7, 7, 2, 3, 8, 7, 7, 3, 8, 9, 1, 6, 2] | 0 | 2 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 4, 5, 6, 6, 5, 4, 6, 6, 6, 6, 6, 6, 5, 3, 2, 4, 3, 1, 7, 5, 6] | 2 | 0 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 2, 1, 2, 3, 4, 3, 2, 1, 1, 1, 6, 2, 1, 1, 1, 2, 1, 2, 3, 1, 2] | 1 | 1 |
| [1, 3, 2, 2, 1, 1, 1, 1, 1, 4, 2, 3, 6, 2, 1, 2, 3, 4, 2, 1, 1, 1] | 1 | 1 |
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [2, 1, 2, 3, 4, 4, 3, 2, 1, 3, 2, 1, 2, 1, 3, 3, 2, 1, 3, 2, 1, 1] | 1 | 1 |
| [1, 2, 3, 2, 1, 3, 2, 1, 2, 7, 6, 2, 2, 2, 2, 3, 2, 1, 2, 3, 2, 3] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 4, 5, 6, 5, 6, 5, 5, 5, 6, 6, 6, 5, 1, 4, 3, 2, 4, 6, 2, 4, 1] | 2 | 0 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 2, 3, 4, 5, 1, 2, 3, 4] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 3, 2, 2, 1, 1, 1, 1, 1, 4, 2, 3, 6, 2, 1, 2, 3, 4, 2, 1, 1, 1] | 1 | 1 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 3, 2, 2, 1, 1, 1, 1, 1, 4, 2, 3, 6, 2, 1, 2, 3, 4, 2, 1, 1, 1] | 1 | 1 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 2, 3, 4, 5, 1, 2, 3, 4] | 1 | 1 |
| [2, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 1, 5] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 5, 4, 6, 5, 5, 3, 2, 1, 4, 7, 2, 1, 6] | 2 | 0 |
| [1, 6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 4, 2, 3, 2, 1, 2, 3, 2, 1, 5, 1, 4, 2, 6, 1, 2, 4, 2, 1, 2, 3] | 1 | 1 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 5, 4, 6, 5, 5, 3, 2, 1, 4, 7, 2, 1, 6] | 2 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 3, 1, 5, 3, 4, 2, 2, 2, 2, 4, 2, 3, 1, 3, 7, 8, 6, 2, 1, 7, 2] | 2 | 1 |
| [1, 3, 1, 1, 1, 2, 3, 4, 1, 3, 2, 4, 3, 2, 5, 2, 1, 2, 3, 4, 5, 2] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 2, 1, 2, 3, 4, 3, 2, 1, 1, 1, 6, 2, 1, 1, 1, 2, 1, 2, 3, 1, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 1, 2, 3, 4, 4, 3, 2, 1, 3, 2, 1, 2, 1, 3, 3, 2, 1, 3, 2, 1, 1] | 1 | 1 |
| [1, 6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 5, 4, 6, 5, 5, 3, 2, 1, 4, 7, 2, 1, 6] | 2 | 0 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 2, 3, 4, 2, 1, 1, 2, 4, 3, 2, 1, 5, 2, 1, 3, 2, 1, 3, 2, 1, 2] | 1 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 2, 3, 4, 2, 1, 1, 2, 4, 3, 2, 1, 5, 2, 1, 3, 2, 1, 3, 2, 1, 2] | 1 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 5, 4, 6, 5, 5, 3, 2, 1, 4, 7, 2, 1, 6] | 2 | 0 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |

| data_test | actual | prediction |
|--|--------|------------|
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 2, 3, 6, 7, 7, 7, 2, 4, 3, 7, 4, 2, 1, 3, 2, 2, 1, 2, 5, 1] | 1 | 0 |
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [2, 1, 2, 3, 4, 4, 3, 2, 1, 3, 2, 1, 2, 1, 3, 3, 2, 1, 3, 2, 1, 1] | 1 | 1 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 2, 1, 2, 3, 4, 3, 2, 1, 1, 1, 6, 2, 1, 1, 1, 2, 1, 2, 3, 1, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 2, 1, 2, 3, 4, 3, 2, 1, 1, 1, 6, 2, 1, 1, 1, 2, 1, 2, 3, 1, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [1, 3, 2, 1, 2, 3, 5, 1, 2, 7, 2, 1, 1, 1, 1, 1, 3, 2, 3, 2, 3, 3] | 1 | 1 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 4, 5, 6, 5, 6, 5, 5, 5, 6, 6, 6, 5, 1, 4, 3, 2, 4, 6, 2, 4, 1] | 2 | 0 |
| [1, 3, 1, 1, 1, 2, 3, 4, 1, 3, 2, 4, 3, 2, 5, 2, 1, 2, 3, 4, 5, 2] | 1 | 1 |
| [2, 2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 2, 1, 2, 3, 4, 3, 2, 1, 1, 1, 6, 2, 1, 1, 1, 2, 1, 2, 3, 1, 2] | 1 | 1 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 1, 5] | 2 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 8, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 4, 5, 6, 5, 6, 5, 5, 5, 6, 6, 6, 5, 1, 4, 3, 2, 4, 6, 2, 4, 1] | 2 | 0 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 3, 2, 1, 2, 3, 5, 1, 2, 7, 2, 1, 1, 1, 1, 1, 3, 2, 3, 2, 3, 3] | 1 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [2, 1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 4, 2, 3, 2, 1, 2, 3, 2, 1, 5, 1, 4, 2, 6, 1, 2, 4, 2, 1, 2, 3] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 4, 5, 6, 5, 6, 5, 5, 5, 6, 6, 6, 5, 1, 4, 3, 2, 4, 6, 2, 4, 1] | 2 | 0 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 2, 1, 2, 1, 2, 1, 2, 3, 2, 4, 2, 1, 3, 4, 5, 1, 3, 2, 1, 2, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |

Model C (70:30)

| data_test | actual | prediction |
|--|--------|------------|
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 4, 5, 6, 6, 5, 4, 6, 6, 6, 6, 6, 5, 3, 2, 4, 3, 1, 7, 5, 6] | 2 | 0 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 2, 3, 2, 1, 3, 2, 1, 2, 7, 6, 2, 2, 2, 2, 3, 2, 1, 2, 3, 2, 3] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 2, 3, 6, 7, 7, 7, 7, 2, 4, 3, 7, 4, 2, 1, 3, 2, 2, 1, 2, 5, 1] | 1 | 0 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [1, 2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |
| [1, 6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [2, 2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |

| data_test | actual | prediction |
|--|--------|------------|
| [2, 3, 1, 4, 2, 3, 2, 3, 3, 2, 2, 4, 2, 2, 2, 3, 4, 1, 5, 2, 6, 2] | 1 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 3, 2, 1, 3, 2, 4, 5, 1, 6, 2, 3, 2, 4, 3, 2, 1, 2, 4, 2, 3, 2] | 1 | 1 |
| [1, 5, 6, 6, 5, 6, 5, 6, 5, 8, 5, 5, 5, 4, 3, 6, 2, 1, 2, 1, 6, 2] | 2 | 0 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [2, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 2, 6, 2, 3, 6, 6, 6, 4, 6, 8, 7, 6, 5, 5, 4, 6, 5, 4, 6, 5, 7] | 0 | 0 |
| [1, 3, 1, 1, 1, 2, 3, 4, 1, 3, 2, 4, 3, 2, 5, 2, 1, 2, 3, 4, 5, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 6, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 1, 2, 3, 4, 4, 3, 2, 1, 3, 2, 1, 2, 1, 3, 3, 2, 1, 3, 2, 1, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 5, 2, 3, 1, 2, 3, 5, 2, 2, 5, 1, 3, 1, 1, 1, 1, 3, 2, 4, 2] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 8, 7, 7, 6, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 2, 1, 2, 1, 2, 1, 2, 3, 2, 4, 2, 1, 3, 4, 5, 1, 3, 2, 1, 2, 2] | 1 | 1 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |

| data_test | actual | prediction |
|--|--------|------------|
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 8, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 1, 2, 4, 5, 6, 5, 5, 4, 6, 4, 7, 2, 3, 8, 7, 3, 8, 3, 2, 3, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 8, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 2, 3, 6, 7, 7, 7, 7, 2, 4, 3, 7, 4, 2, 1, 3, 2, 2, 1, 2, 5, 1] | 1 | 0 |
| [1, 5, 6, 6, 5, 6, 5, 6, 5, 8, 5, 5, 5, 4, 3, 6, 2, 1, 2, 1, 6, 2] | 2 | 0 |
| [1, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 1, 2, 3, 4, 4, 3, 2, 1, 3, 2, 1, 2, 1, 3, 3, 2, 1, 3, 2, 1, 1] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |
| [2, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 1, 5] | 2 | 1 |
| [2, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 2, 3, 1, 3, 2, 5, 2, 1, 2, 1, 2, 5, 3, 2, 1, 2, 3, 1, 3, 4, 2] | 1 | 1 |
| [1, 2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 2, 1, 2, 3, 4, 3, 2, 1, 1, 1, 6, 2, 1, 1, 1, 2, 1, 2, 3, 1, 2] | 1 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [2, 2, 1, 2, 1, 2, 1, 2, 3, 2, 4, 2, 1, 3, 4, 5, 1, 3, 2, 1, 2, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 8, 7, 7, 3, 8, 9, 1, 6, 2] | 0 | 2 |
| [2, 3, 2, 1, 3, 2, 4, 5, 1, 6, 2, 3, 2, 4, 3, 2, 1, 2, 4, 2, 3, 2] | 1 | 1 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 2, 7, 6, 7, 6, 7, 2, 3, 1] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 4, 2, 3, 2, 1, 2, 3, 2, 1, 5, 1, 4, 2, 6, 1, 2, 4, 2, 1, 2, 3] | 1 | 1 |
| [2, 2, 6, 2, 3, 6, 6, 6, 4, 6, 8, 7, 6, 5, 5, 4, 6, 5, 4, 6, 5, 7] | 0 | 0 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [1, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [2, 5, 2, 3, 1, 2, 3, 5, 2, 2, 5, 1, 3, 1, 1, 1, 1, 1, 3, 2, 4, 2] | 1 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 3, 2, 1, 3, 2, 4, 5, 1, 6, 2, 3, 2, 4, 3, 2, 1, 2, 4, 2, 3, 2] | 1 | 1 |
| [2, 3, 2, 2, 1, 1, 1, 1, 1, 4, 2, 3, 6, 2, 1, 2, 3, 4, 2, 1, 1, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 2, 3, 2, 1, 3, 2, 1, 2, 7, 6, 2, 2, 2, 3, 2, 1, 2, 3, 2, 3] | 1 | 1 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 2, 3, 5, 2, 1, 1, 1, 1, 4, 3, 2, 4, 2, 1, 3, 4, 6, 1, 3, 2, 2] | 1 | 1 |
| [1, 2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 2, 3, 4, 5, 1, 2, 3, 4] | 1 | 1 |

| data_test | actual | prediction |
|--|--------|------------|
| [2, 5, 2, 3, 1, 2, 3, 5, 2, 2, 5, 1, 3, 1, 1, 1, 1, 1, 3, 2, 4, 2] | 1 | 1 |
| [2, 2, 6, 2, 3, 6, 6, 6, 4, 6, 8, 7, 6, 5, 5, 4, 6, 5, 4, 6, 5, 7] | 0 | 0 |
| [2, 4, 2, 3, 4, 2, 3, 2, 4, 2, 4, 2, 4, 1, 3, 4, 2, 4, 2, 4, 3, 1] | 1 | 1 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 2, 6, 2, 3, 6, 6, 6, 4, 6, 8, 7, 6, 5, 5, 4, 6, 5, 4, 6, 5, 7] | 0 | 0 |
| [1, 6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [2, 3, 2, 1, 3, 2, 1, 3, 2, 1, 2, 2, 2, 2, 2, 1, 3, 4, 4, 1, 4, 1] | 1 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 5, 4, 6, 5, 5, 3, 2, 1, 4, 7, 2, 1, 6] | 2 | 0 |
| [1, 4, 2, 3, 2, 1, 2, 3, 2, 1, 5, 1, 4, 2, 6, 1, 2, 4, 2, 1, 2, 3] | 1 | 1 |
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 1, 1, 2, 3, 2, 5, 1, 5, 1, 1] | 1 | 1 |
| [1, 4, 5, 6, 5, 6, 5, 5, 5, 6, 6, 6, 5, 1, 4, 3, 2, 4, 6, 2, 4, 1] | 2 | 0 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 3, 1, 5, 3, 4, 2, 2, 2, 2, 4, 2, 3, 1, 3, 7, 8, 6, 2, 1, 7, 2] | 2 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 9, 6, 5, 7, 2, 4, 3, 1, 4] | 0 | 0 |
| [1, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 5, 7, 6, 7, 8, 7, 6, 2] | 0 | 0 |
| [1, 6, 7, 8, 7, 6, 7, 7, 2, 4, 3, 7, 8, 2, 3, 6, 4, 2, 3, 1, 2, 1] | 1 | 0 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 8, 7, 7, 3, 8, 9, 1, 6, 2] | 0 | 2 |
| [1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 1, 5] | 2 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 6, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |

| data_test | actual | prediction |
|--|--------|------------|
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [1, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [2, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 8, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [1, 4, 5, 6, 6, 5, 4, 6, 6, 6, 6, 6, 5, 3, 2, 4, 3, 1, 7, 5, 6] | 2 | 0 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [2, 1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 8, 7, 7, 5, 3, 2, 7, 8, 2, 4, 5, 3] | 0 | 0 |
| [2, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |
| [2, 4, 2, 3, 4, 2, 3, 2, 4, 2, 4, 2, 4, 1, 3, 4, 2, 4, 2, 4, 3, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 3, 2, 4, 2, 3, 2, 3, 3, 2, 2, 3, 3, 4, 5, 6, 5, 5, 4, 6, 5, 4] | 2 | 1 |
| [2, 1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [1, 4, 5, 6, 6, 5, 4, 6, 6, 6, 6, 6, 5, 3, 2, 4, 3, 1, 7, 5, 6] | 2 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 4, 8, 7, 7, 4, 4, 5, 6, 5, 5, 4, 6, 5] | 0 | 0 |
| [1, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 2, 3, 4, 2, 4, 3, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 1, 5] | 2 | 1 |
| [2, 1, 2, 3, 4, 2, 4, 3, 3, 2, 3, 4, 4, 1, 2, 4, 6, 5, 4, 2, 5] | 2 | 1 |
| [2, 1, 2, 2, 3, 2, 4, 2, 3, 2, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 2, 1] | 1 | 1 |
| [1, 6, 8, 7, 7, 7, 6, 2, 4, 1, 2, 4, 3, 2, 7, 6, 5, 1, 9, 3, 4, 2] | 2 | 2 |
| [1, 2, 4, 5, 4, 3, 2, 2, 4, 3, 2, 2, 4, 3, 4, 2, 2, 3, 1, 2, 3, 4] | 1 | 1 |
| [2, 1, 6, 7, 8, 7, 6, 7, 7, 3, 4, 8, 7, 3, 2, 6, 4, 2, 3, 1, 2, 1] | 2 | 0 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |
| [2, 6, 8, 7, 7, 7, 6, 7, 7, 8, 7, 7, 9, 3, 2, 4, 1, 4, 2, 4, 2, 3] | 0 | 0 |
| [2, 4, 2, 3, 4, 2, 3, 2, 4, 2, 4, 2, 4, 1, 3, 4, 2, 4, 2, 4, 3, 1] | 1 | 1 |

| data_test | actual | predication |
|--|--------|-------------|
| [2, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 4, 5, 6, 5, 5, 4, 6, 7, 2, 3, 4, 8, 8, 7, 9, 2, 1, 4, 6, 7, 2] | 0 | 2 |
| [1, 3, 1, 1, 1, 2, 3, 4, 1, 3, 2, 4, 3, 2, 5, 2, 1, 2, 3, 4, 5, 2] | 1 | 1 |
| [1, 2, 3, 2, 1, 3, 2, 1, 2, 7, 6, 2, 2, 2, 2, 3, 2, 1, 2, 3, 2, 3] | 1 | 1 |
| [1, 2, 4, 5, 6, 5, 5, 4, 6, 5, 4, 6, 5, 5, 3, 2, 1, 4, 7, 2, 1, 6] | 2 | 0 |
| [1, 2, 1, 5, 3, 2, 3, 2, 4, 1, 4, 2, 4, 6, 7, 2, 5, 8, 1, 3, 2, 3] | 2 | 1 |
| [1, 6, 8, 7, 7, 6, 7, 7, 3, 8, 7, 9, 6, 5, 2, 5, 2, 3, 2, 1, 7, 6] | 0 | 0 |
| [2, 1, 6, 7, 5, 3, 2, 6, 2, 3, 3, 2, 2, 3, 3, 7, 7, 4, 8, 7, 7, 5] | 2 | 2 |
| [1, 3, 1, 1, 1, 2, 3, 4, 1, 3, 2, 4, 3, 2, 5, 2, 1, 2, 3, 4, 5, 2] | 1 | 1 |
| [1, 1, 3, 2, 4, 2, 6, 2, 2, 2, 1, 3, 4, 4, 2, 2, 2, 3, 2, 1, 2, 4] | 1 | 1 |
| [1, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 8, 4, 2, 3, 1, 4, 5, 6, 7, 5] | 0 | 0 |
| [2, 3, 1, 4, 3, 2, 3, 4, 3, 1, 4, 3, 1, 3, 2, 2, 4, 2, 2, 3, 4, 3] | 1 | 1 |
| [2, 2, 1, 3, 4, 3, 5, 3, 2, 6, 1, 1, 2, 3, 2, 1, 3, 2, 4, 2, 1, 1] | 1 | 1 |