Klasifikasi Data Kesehatan di Indonesia

Daniel Davinci1, Dapot H. Y Sidabutar2, Faikar Fawwaz Muafif3, Anju Wahyu4

Program Studi Sistem Informasi, Universitas Komputer Indonesia   
Bandung

[1daniel.10519192@mahasiswa.unikom.ac.id](mailto:1daniel.10519192@mahasiswa.unikom.ac.id)

[2dapot.10519187@mahasiswa.unikom.ac.id](mailto:2dapot.10519187@mahasiswa.unikom.ac.id)

[3](mailto:3faikar.10519165@mahasiswa.unikom.ac.id)[faikar.10519165](mailto:3faikar.10519165@mahasiswa.unikom.ac.id)[@mahasiswa.unikom.ac.id](mailto:3faikar.10519165@mahasiswa.unikom.ac.id)

[4](mailto:4anju.10519172@mahasiswa.unikom.ac.id)[anju.10519172@mahasiswa.unikom.ac.id](mailto:4anju.10519172@mahasiswa.unikom.ac.id)

Abstrak— *Kesehatan adalah hal yang harus diperhatikan dalam kehidupan sehari-hari. Tingkat kesehatan pada suatu negara juga berkaitan dengan produktifitas negara tersebut. Di negara Indonesia ini tingkat kesehatan masih terggolong rendah karena masih banyaknya ketidaksadaran masyarakat dalam menjaga kesehatan dirinya sendiri. Di samping hal itu juga adanya wabah yang melanda negara Indonesia seperti COVID-19 pada tahun 2019 juga sangat berdampak besar bagi kesehatan di Indonesia. Untuk mengetahui rata-rata kesehatan yang berada di Indonesia ini maka akan dilakukannya klasifikasi data kesehatan yang berada di Indonesia menggunakan metode data mining. Data yang terkait adalah data yang berhubungan dengan perkembangan kesehatan di Indonesia, baik data pasien hingga data tingkat kematian di Indonesia.*

***Kata kunci***— **kesehatan, Indonesia, COVID-19**

***Abstract- Health is something that must be considered in everyday life. The level of health in a country is also related to the productivity of that country. In the country of Indonesia, the level of health is still classified as low because there is still a lot of public unawareness in maintaining their own health. In addition to this, the outbreaks that hit the country of Indonesia such as Covid-19 in 2019 also had a huge impact on health in Indonesia. To find out the average health in Indonesia, the classification of health data in Indonesia will be carried out using data mining methods. The related data is data related to health development in Indonesia, both patient data to data on mortality rates in Indonesia.***

Key word— Health, Indoensia, COVID-19

1. **Pendahuluan**

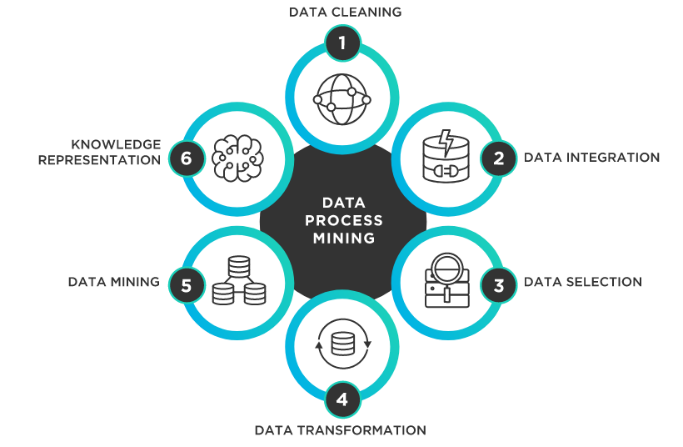
Seiring perkembangan zaman teknologi juga mengalami perkembangan yang signifikan dan dapat di kategorikan memiliki perkembangan yang pesat. Perkembangan teknologi ini sangat membantu manusia untuk mempermudia menyelesaikan pekerjaannya. Salah satu perkembangan dari teknologi sendiri adalah kemajuannya komputer yang semakin canggih dan semakin mampu menangani data yang semakin banyak lagi dalam pengolahannya. Seiring dengan perkembangan komputer, tentu saja metode pengolahan data yang membutuhkan komputer sebagai perantara pengolahan datanya juga semakin berkembang. Salah satu metode pengolahan data yang sudah terkenal dan banyak dipakai orang-orang dalam mengolah data adalah metode data mining.

Dengan adanya teknologi yang semakin canggih ini akan sangat membantu berbagai bidang dalam pengolahan data yang besar. Salah satu bidang yang membutuhkan pengolahan data yang besar sendiri adalah bidang kesehatan. Seperti yang kita tahu, pada tahun 2019 terdapat wabah yang bernama COVID-19 yang menyerang dunia dalam kesehatan. Di Indonesia sendiri terdapat banyak korban yang berjatuhkan yang diakibatkan oleh wabah COVID-19 itu sendiri. Terdapat banyak data juga yang membahas tingkat kesehatan di Indonesia ini sejak terjadinya wabah tersebut.

Dalam pengolahan data menggunakan data mining ini tentu saja memerlukan data untuk diolah. Pada pengolahan data mining kali ini kelompok kami memilih untuk melakukan pengolahan data dengan menggunakan data yang membahas mengenai kesehatan yang berada di Indonesia. Metode data mining yang akan kami gunakan dalam pengolahan data-data yang kami miliki ini adalah K-Means dan Naïve Bayes.

1. **Tinjauan pustaka**
   1. **Data Mining**

Pengertian data mining sendiri adalah suatu proses pengolahan data yang menggunakan machine learning untuk pengolahan datanya. Dalam pengolahan data menggunakan data mining juga terdapat juga tahapan untuk mencapai hasil akhirnya, tahapan tersebut adalah sebagai berikut



1. Data Cleaning

Data Cleaning ini adalah proses untuk menghilangkan suatu data yang tidak berhubungan atau relevean dengan data lainnya.

1. Data Intergration

Data Intergration ini adalah penggabungan dari suatu data yang berasal dari berbagai data ke data barunya.

1. Data Selection

Data Selection ini adalah data yang terdapat pada basis data namun biasanya tidak di pakai secara keseluruhan maka diadakannya penyeleksian data.

1. Data Transformation

Data Transformation adalah data yang diubah dan akan mengalami perubahan setelah prosesnya.

1. Mining

Proses mining adalah proses yang utama untuk mengetahui informasi yang penting dari suatu data.

1. Pattern Evaluation

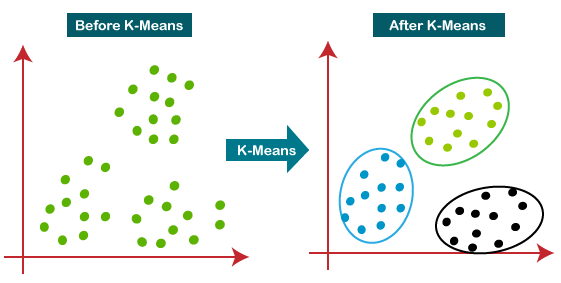
Pattern Evaluation adalah mengevaluasi suatu pola yang bertujuan untuk menemukan berbagai macam pola lainnya pada suatu data.

1. Knowledge Presentation

Knowledge Presentation ini adalah sebuah pengetahuan mengenai penyajian dari data yang telah diolah.

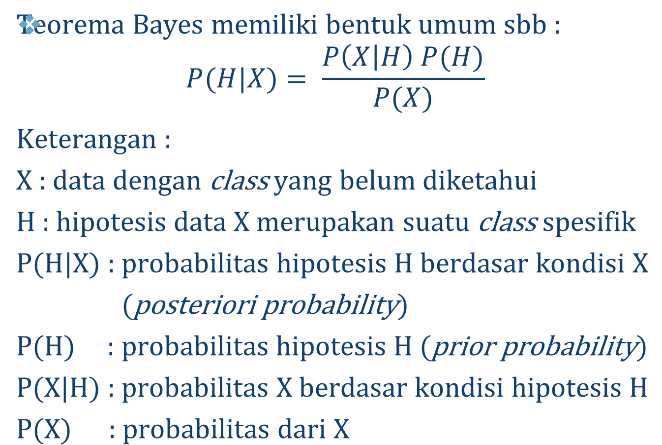
* 1. **K-Means**

K-means sendiri memilik definisi sebagai algoritma yang biasanya digunakan untuk sebuah clustering pada suatu data. Metode K-means sendiri sangat membantu untuk user dalam mengelompokkan data ke dalam grup yang berbeda berdasarkan dari variabelnya.



* 1. **Naïve Bayes**

Naïve Bayes sendiri adalah sebagai salah satu metode klasifikasi data yang sangat sering digunakan banyak orang. Metode ini sangat sering dipakai banyak orang karena metode ini memiliki hasil yang tingkat akuratnya dapat di kategorikan sebagai kategori yang baik. Metode ini juga memiliki sebuah rumus yang harus digunakan dalam pelaksanaannya, yaitu adalah sebagai berikut



1. **pembahasan**
   1. **Metode**

Pada pengolahan data ini kelompok kami memutuskan untuk menggunakan pengolahan data dengan metode K-Means dan Naïve Bayes.

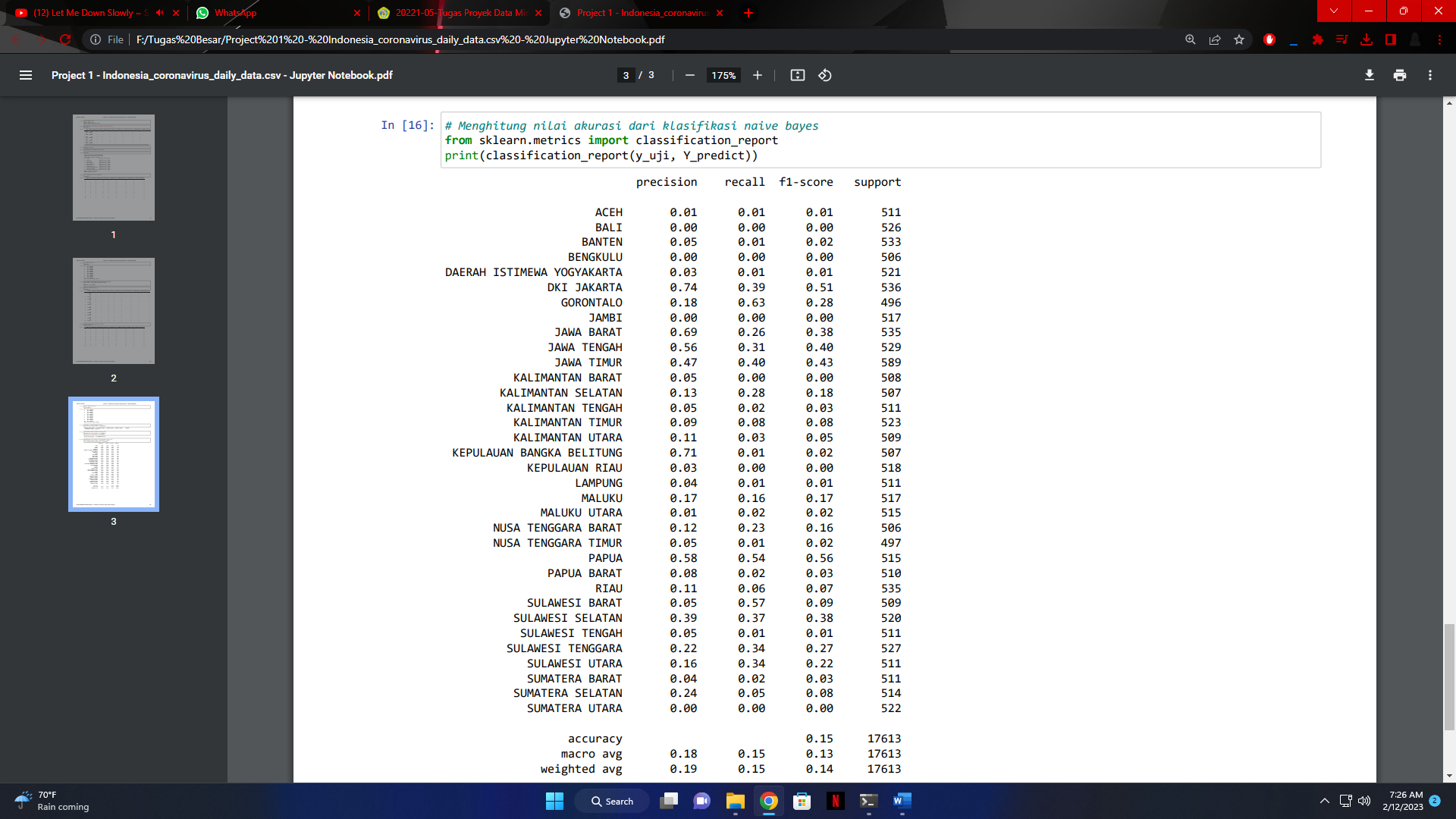
* 1. **Data**

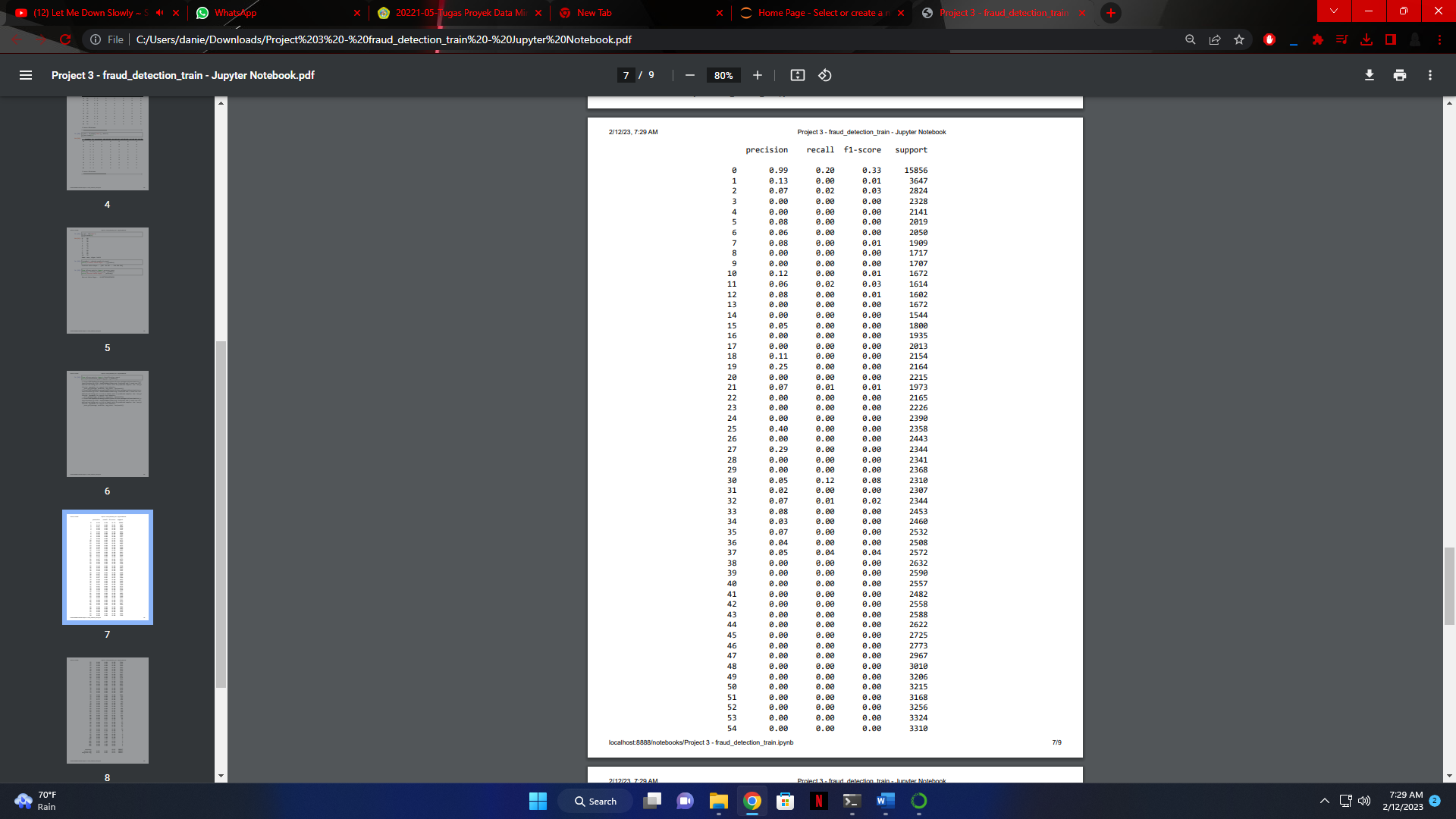
Data-data yang digunakan dalam proses data mining yang kelompok kami lakukan ini adalah data yang berhubungan dengan kesehatan yang berada di Indonesia. Data-data mengenai kesehatan ini kami ambil dari salah satu situs data resmi yaitu Kaggle.

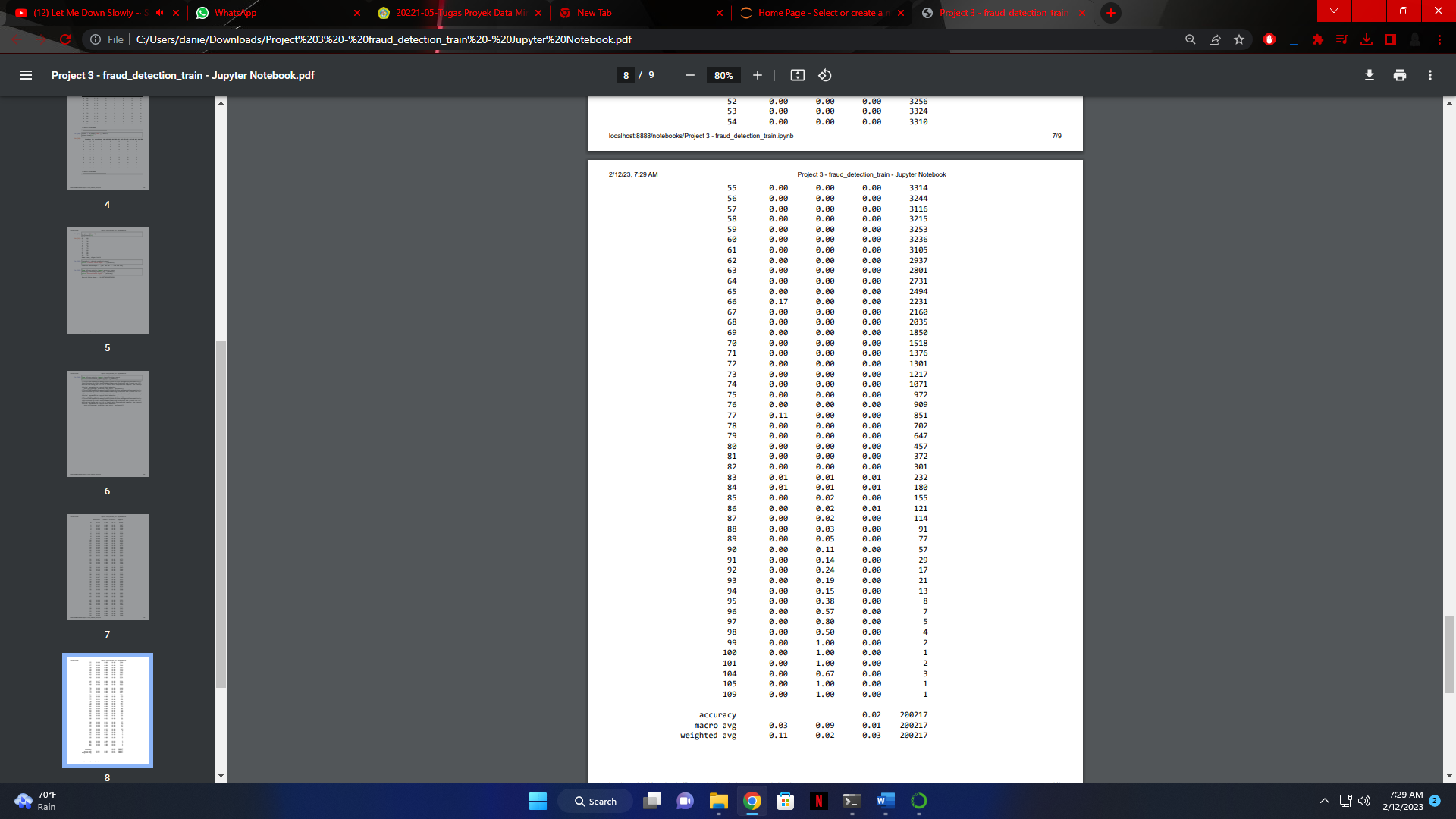
Berikut adalah data yang terkait yang kami gunakan dalam pengolahan data mining ini

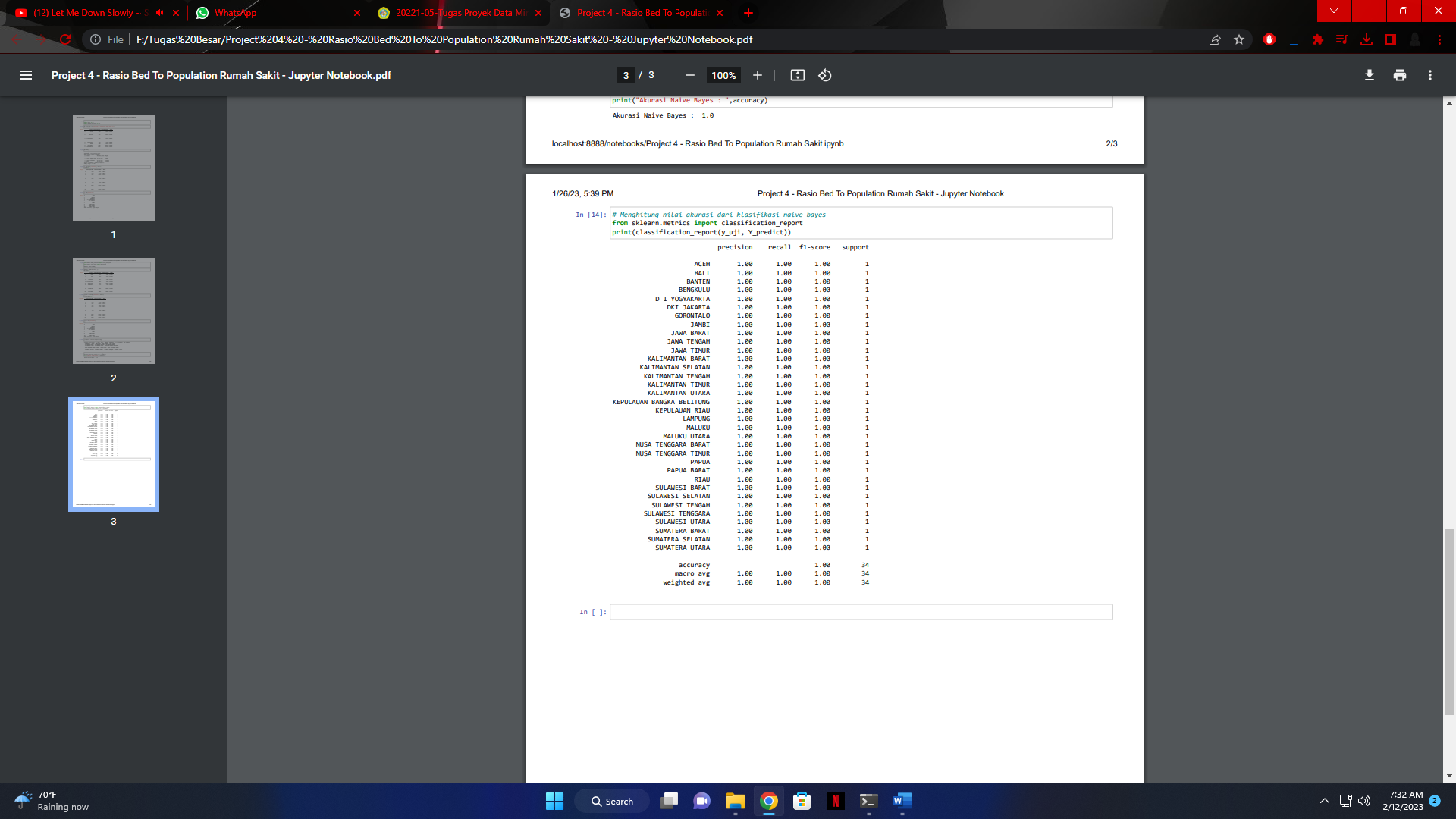
1. COVID-19 Data Set
2. COVID-19 Indonesia Dataset (Case and Vaccination)
3. BPJS Kesehatan Hackathon 2021 - Cost Prediction
4. List Faskes BPJS Indonesia
5. Indonesian Abusive and Hate Speech Twitter Text
6. COVID-19 Indonesian Tweets
7. Causes of Death in Indonesia
8. Word library kesehatan indonesia
9. Dataset Rasio Bed to Population Faskes II
10. BPJS Kesehatan Hackathon 2021 - Fraud Detection
    1. **Hasil**

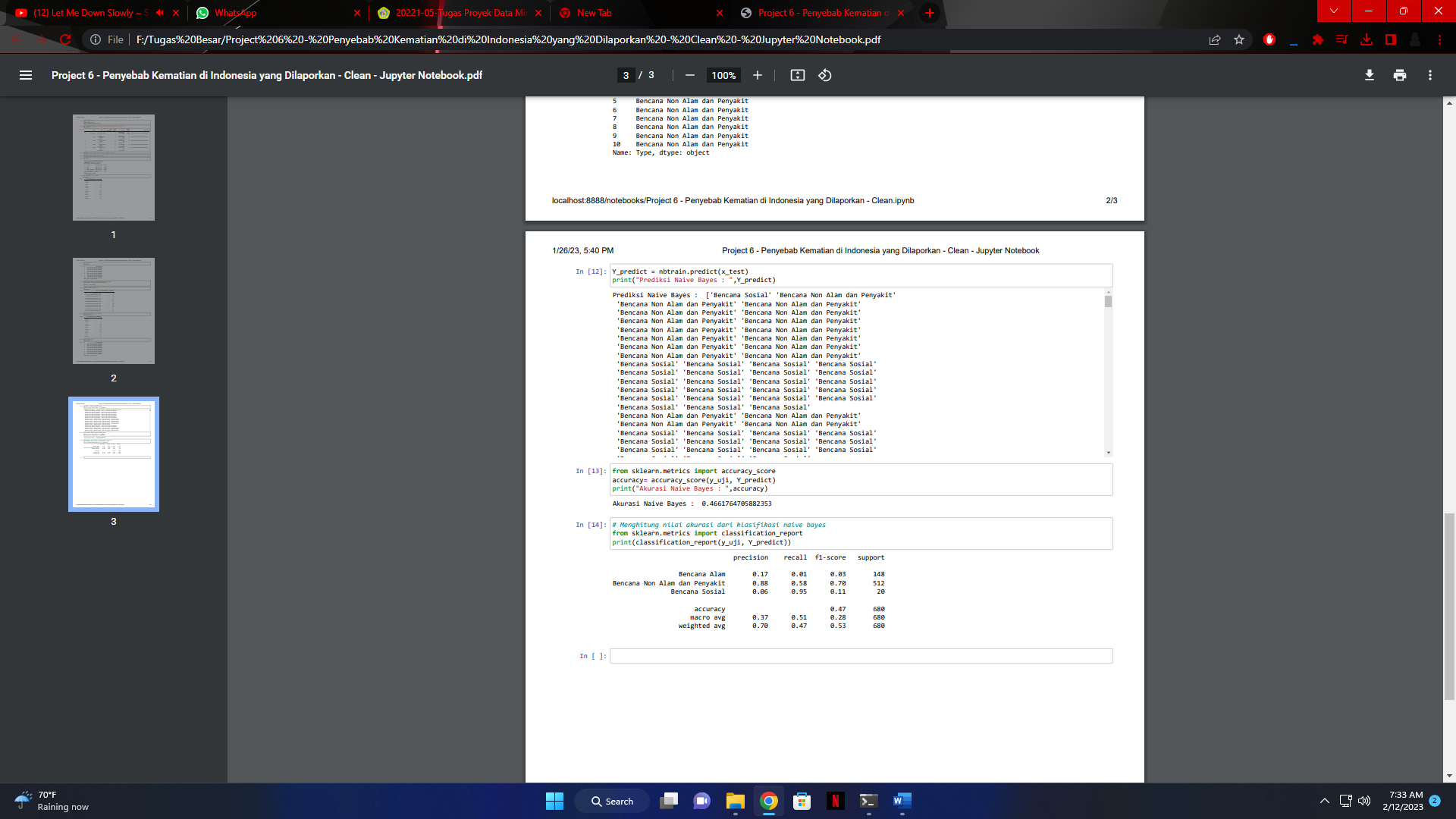
Setelah melakukan pengumpulan data yang bersumber dari Kaggle, kami mengolah data-data tersebut dengan data mining menggunakan metode Naïve Bayes dan K-Means. Berikut adalah beberapa sample dari hasil akhir yang telah kami kerjakan :

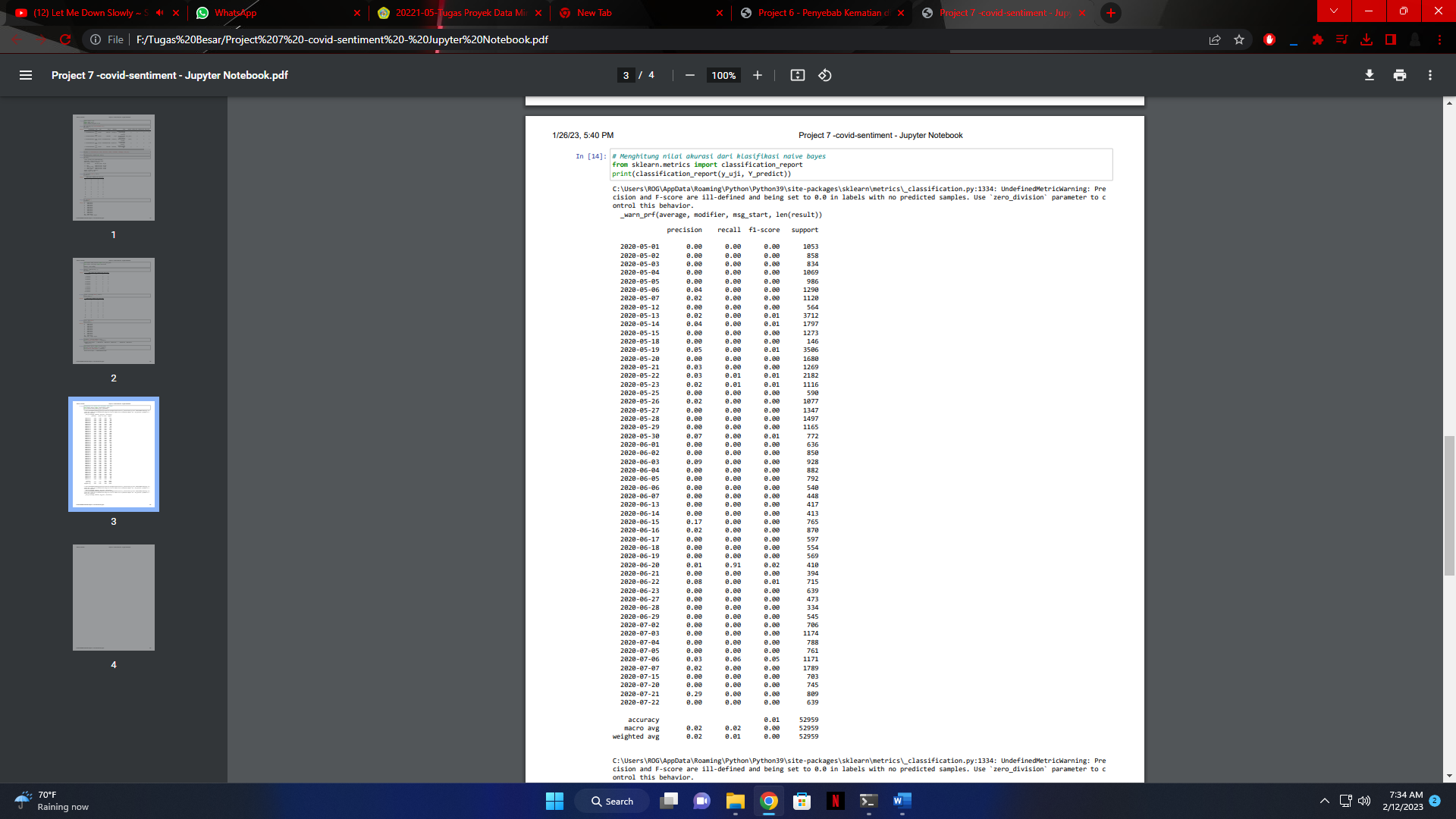


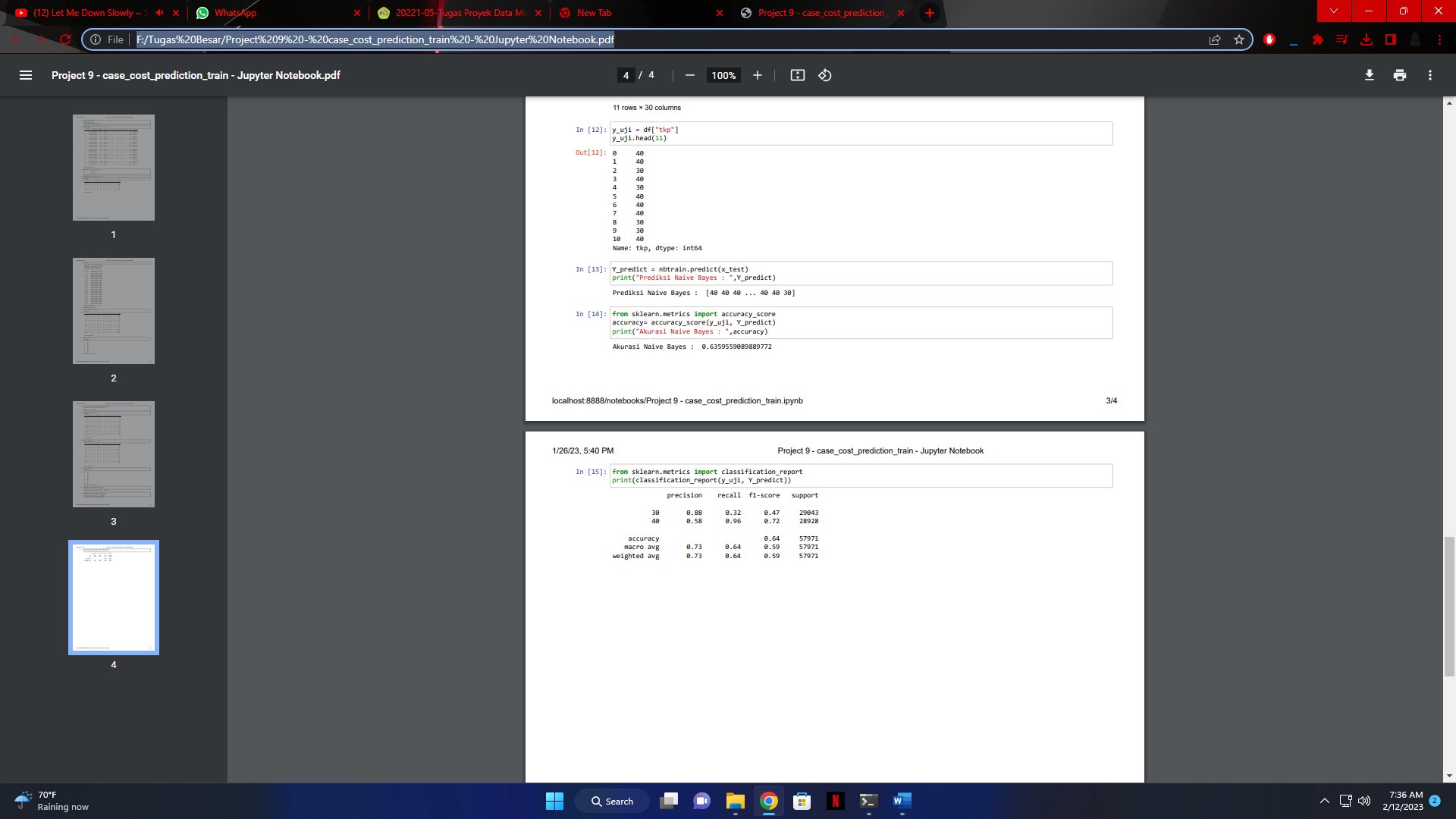


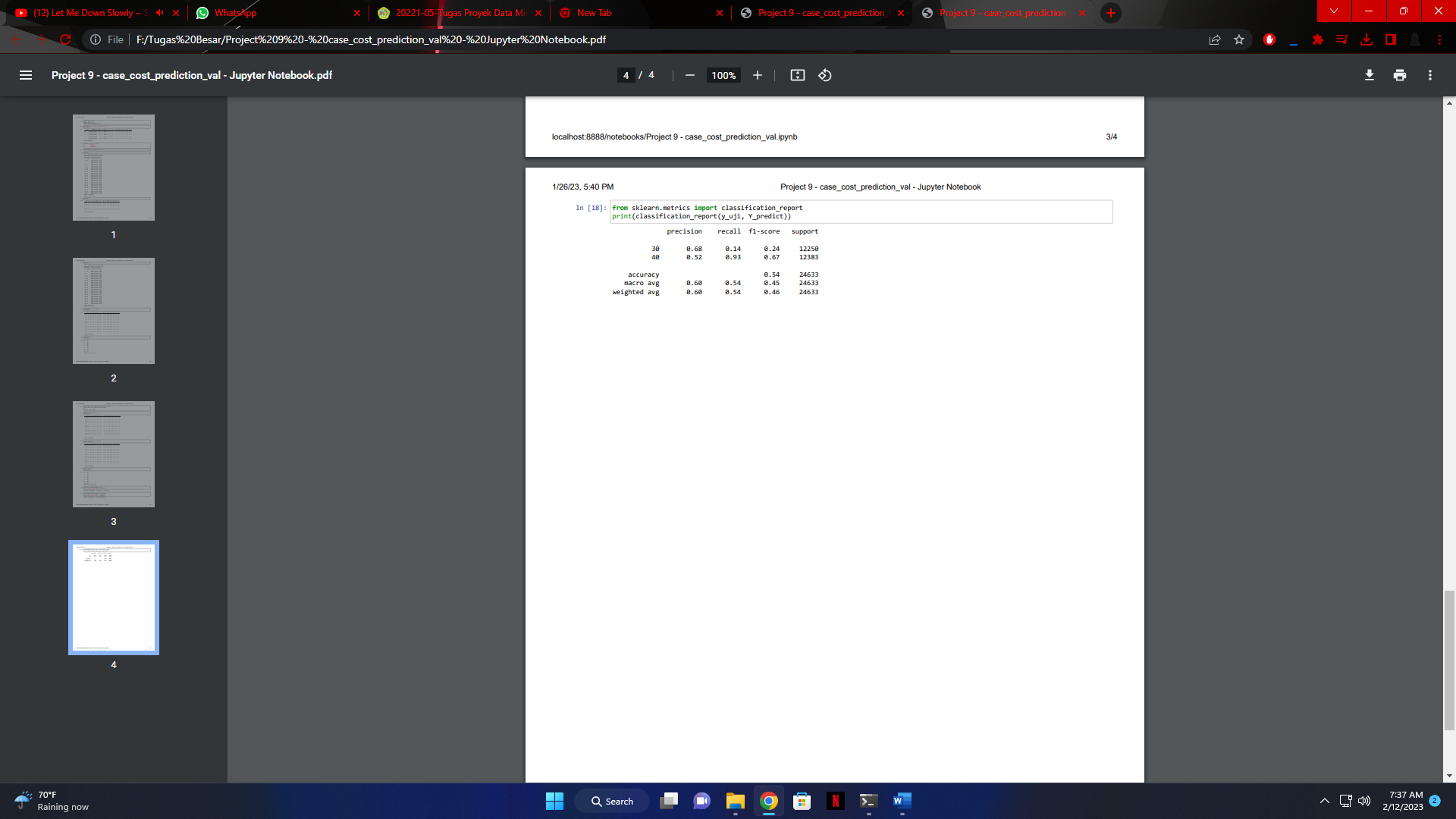


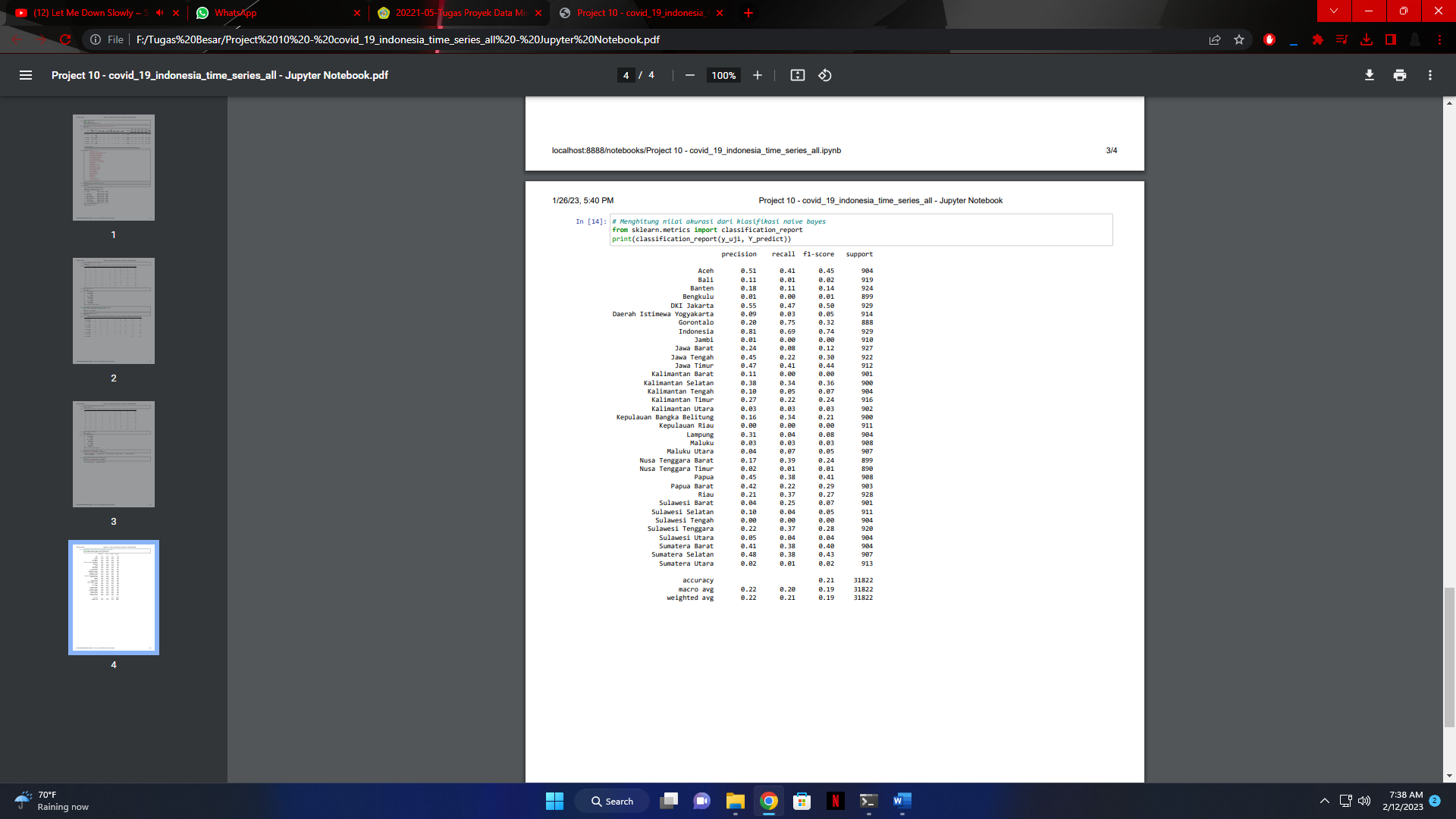






****

****



1. **Kesimpulan**

Berdasarkan pengolahan data yang telah kelompok kami lakukan. Pengolahan data ini berhasil mengolah data-data yang sudah kami siapkan sebelumnya dan menunjukan hasil yang sesuai dengan metode yang kami gunakan, yaitu metode K-Means dan metode Naïve Bayes. Dari hasil yang telah kami dapatkan, data-data tersebut terolah dengan akurat dan tepat sesuai dengan pengelompokkan datanya masing-masing.

1. **Daftar pustaka**

|  |  |
| --- | --- |
| [1] | F. Afshoh, *Analisa Sentimen Menggunakan Naive Bayes,* (2017). |
| [2] | Kaggle. [Online]. Available: <https://www.kaggle.com/> . |
| [3] | Hendratno, “covid19-indonesia,” [Online]. Available: <https://www.kaggle.com/datasets/hendratno/covid19-indonesia> . |
| [4] | Bagusbpg, “bpjs-kesehatan-hackathon-2021-cost-prediction,” [Online]. Available: <https://www.kaggle.com/datasets/bagusbpg/bpjs-kesehatan-hackathon-2021-cost-prediction> |
| [5] | Riqulaziz, “case-vaccination-covid19-indonesia-dataset,” [Online]. Available: <https://www.kaggle.com/datasets/riqulaziz/case-vaccination-covid19-indonesia-dataset>. |
| [6] | Israhabibi, “list-faskes-bpjs-indonesia,” [Online]. Available: <https://www.kaggle.com/datasets/israhabibi/list-faskes-bpjs-indonesia> |
| [7] | Ilhamfp31, “indonesian-abusive-and-hate-speech-twitter-text,” [Online]. Available: <https://www.kaggle.com/datasets/ilhamfp31/indonesian-abusive-and-hate-speech-twitter-text> |
| [8] | Dionisiusdh, “covid19-indonesian-twitter-sentiment,” [Online]. Available: <https://www.kaggle.com/datasets/dionisiusdh/covid19-indonesian-twitter-sentiment>. |
| [9] | Hendratno, “cause-of-death-in-indonesia,” [Online]. Available: <https://www.kaggle.com/datasets/hendratno/cause-of-death-in-indonesia>. |
| [10] | Alvinf, “word-library-kesehatan-indonesia,” [Online]. Available: <https://www.kaggle.com/datasets/alvinf/word-library-kesehatan-indonesia> |
| [11] | Yafethtb, “dataset-rasio-bed-to-population-faskes-ii,” [Online]. Available: <https://www.kaggle.com/datasets/yafethtb/dataset-rasio-bed-to-population-faskes-ii> |
| [12] | Bagusbpg, “bpjs-kesehatan-hackathon-2021-fraud-detection,” [Online]. Available: <https://www.kaggle.com/datasets/bagusbpg/bpjs-kesehatan-hackathon-2021-fraud-detection> |