

Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode *Naive Bayes*

Nurul Alfiah

Prodi Sistem Informasi, STMIK Komputama Majenang
Jl. Raya Majenang - Cimanggu KM 08 No 99 Desa Cilempuyang, Kec. Cimanggu,
Kabupaten Cilacap, Jawa Tengah 53256
vey.alfi@gmail.com

INTISARI

PKH (Program Keluarga Harapan) merupakan sebuah program bantuan sosial bersyarat yang dikeluarkan oleh pemerintah dalam upaya untuk mengentaskan kemiskinan. Dalam implementasinya di wilayah kecamatan Cimanggu masih belum optimal dikarenakan PKH ini banyak yang belum tepat sasaran dalam menentukan siapa penerima PKH yang berhak untuk mendapatkannya. Konsep Data Mining akan mempermudah dalam mengatasi masalah tersebut. Dengan menggunakan Algoritma *Naive Bayes Classification* dan metode CRISP-DM (Cross Industry Standard Process for Data Mining) penelitian ini diharapkan dapat membantu pemerintah khususnya Dinas Sosial Kabupaten Cilacap dalam menentukan siapa penerima PKH yang berhak untuk mendapatkannya.

Kata Kunci – Data Mining, Program Keluarga Harapan (PKH), *Naive Bayes*, CRISP-DM

ABSTRACT

PKH (Program Keluarga Harapan) is a conditional social assistance program issued by the government in an effort to alleviate poverty. In its implementation in the Cimanggu subdistrict, it is still not optimal because many PKH have not been on target in determining who PKH recipients are entitled to get it. The concept of Data Mining will make it easier to solve this problem. By using the *Naive Bayes Classification Algorithm* and the CRISP-DM (Cross Industry Standard Process for Data Mining) method, this research is expected to help the government, especially the Cilacap Regency Social Service, in determining who PKH recipients are entitled to get it.

Keywords – Data Mining, Program Keluarga Harapan (PKH), *Naive Bayes*, CRISP-DM

1. PENDAHULUAN

Data Mining merupakan serangkaian proses dengan memanfaatkan kumpulan data untuk mendapatkan berbagai informasi yang berharga dengan ukuran data yang cukup besar melalui proses penggalian data atau penyaringan data. Semua bidang yang mempunyai sejumlah data besar dapat menerapkan *Data Mining*[1]. Hasil dari pengolahan *Data Mining* tersebut dapat digunakan untuk mengambil informasi dan keputusan di masa yang akan datang. Selain itu ada pengertian lain mengenai *Data Mining* menurut Han J Kamber yaitu *Data Mining* adalah sebuah disiplin ilmu yang mempelajari Metode yang digunakan untuk menemukan pola yang tersembunyi dari *big data* [2]. *Data Mining* ini dapat mengubah kumpulan data yang besar/banyak menjadi pengetahuan/informasi yang diperlukan. Selain itu juga dengan *Data Mining*, pencarian pola baru atau trend baru pada suatu data yang besar dapat dilakukan dengan

mudah sehingga dapat membantu mengambil keputusan atau memprediksi data tertentu dan menganalisis apa yang harus dilakukan.

Ada begitu banyak program pemerintah dari berbagai bidang sebagai bentuk usaha untuk mengentaskan kemiskinan. Program Bantuan Pemerintah ini terbagi untuk Individu, keluarga, dan kelompok kurang mampu dalam berbagai bidang diantaranya : bidang Pangan, Pendidikan, Kesehatan, Energi, Sosial dan Ekonomi, Perumahan, Pertanian, Kelautan dan Perikanan. [3]

Kecamatan Cimanggu merupakan salah satu Kecamatan yang ada di Kabupaten Cilacap, Jawa Tengah dengan luas wilayah sekitar 167,44 KM². Pada tahun 2018 jumlah penduduk yang ada di Kecamatan Cimanggu sebanyak 97.122 jiwa. Kecamatan Cimanggu berbatasan langsung dengan Kecamatan Cipari di sebelah Selatan, Kecamatan Brebes di sebelah Utara, Kecamatan Karangpucung disebelah Timur dan Kecamatan Majenang di sebelah Barat. Kecamatan Cimanggu

memiliki 15 Desa/kelurahan. Dengan pembagian berdasarkan jenis kelamin laki-laki sebanyak 48.740 jiwa dan perempuan 48.383 jiwa.

Dengan jumlah Penduduk yang begitu banyak di Kecamatan Cimanggu, diperlukan analisis yang mendalam untuk menentukan warga tidak mampu yang berhak untuk mendapatkan bantuan sosial bersyarat dari pemerintah. Data penduduk yang ada akan dilakukan survei terlebih dahulu, kemudian dianalisis dan hasil analisis tersebut digunakan untuk menentukan warga miskin yang berhak mendapatkan bantuan sosial bersyarat PKH, tetapi pada kenyataannya dalam penyaluran bantuan PKH banyak/ada beberapa yang tidak tepat sasaran. Hal tersebut disebabkan banyak penduduk miskin yang tidak menerima bantuan PKH salah satunya karena penentuan status keluarga miskin yang berhak menerima bantuan belum optimal. Dengan demikian pada penelitian ini penulis akan menggunakan Algoritme Klasifikasi *Data Mining* yaitu *Naive Bayes Classifier* untuk memprediksi warga tidak mampu di kecamatan cimanggu yang berhak untuk mendapatkan bantuan PKH dari pemerintah. Hasil dari penelitian ini diharapkan dapat menjadi acuan bagi pemerintah daerah khususnya dinas sosial yang menangani PKH dalam menanggulangi permasalahan penerima bantuan PKH di Kecamatan Cimanggu.

Beberapa Penelitian terdahulu yang sudah pernah dilakukan diantaranya : penelitian yang dilakukan [4] dengan menghasilkan kesimpulan Algoritme *Naive Bayes* sangat cocok diterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga memudahkan pengurus desa dalam menentukan keluarga yang berhak menerima beras rastra.

Selain itu, dalam jurnal penelitian yang berjudul [5] dengan kesimpulan bahwa algoritma *K-Means*, *dataset* dengan atribut data yang dimodifikasi keseluruhan memiliki hasil *Cluster* yang lebih baik. Hal ini dikarenakan pada algoritme *K-Means* yang menjadi pusat clusternya adalah *mean* (rata-rata), juga *Euclidean* sebagai fungsi jarak untuk menghitung jarak kedekatan antar objek dengan *centroidnya*.

Penelitian dengan judul [6] Klasifikasi Masyarakat Miskin menggunakan Metode *Naive Bayes* dengan kesimpulan bahwa hasil pengujian *Confussion matrix* dengan teknik split validasi, penggunaan Metode *naive bayes*

terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar atau termasuk dalam kategori *Good*.

Penelitian yang dilakukan oleh [7] menyimpulkan bahwa dengan melihat pedoman pola aturan (*rule*) yang terbentuk maka dilakukan perbandingan dengan kondisi rumah ibadah yang dinilai, sehingga tim seleksi dapat memutuskan pemohon layak atau tidak layak diberikan bantuan dana hibah fasilitas rumah ibadah.

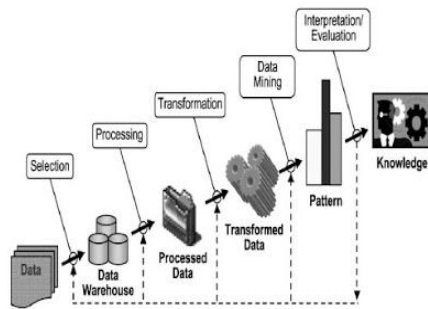
Penelitian yang dilakukan oleh [8] menyimpulkan bahwa hasil komparasi antara algoritma dan *Naive Bayes* untuk mengklasifikasikan tingkat kemiskinan dengan 14 atribut dan jumlah data yang telah di *Cleaning* sebesar 13.928 dataset dapat disimpulkan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih baik dibandingkan dengan Metode *Naive Bayes*.

1.1. METODE PENELITIAN

Salah satu tugas dalam *Data Mining* yaitu memetakan data yang sudah didefinisikan sebelumnya ke dalam satu atau beberapa kelas. Salah satu Metode yang sering digunakan dalam melakukan klasifikasi adalah Metode *Naive Bayes* yang merupakan sebuah Metode pengelompokan statistik pada produk probabilistik sederhana yang digunakan untuk memprediksi peluang atau kemungkinan dengan menjumlahkan kombinasi dan frekuensi dari nilai dataset yang diberikan. Algoritme yang menggunakan teorema Bayes, dapat mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas[9]. *Naive Bayes* memiliki akurasi dan kecepatan yang kuat Ketika diaplikasikan pada database dengan jumlah data yang besar/*big data*. [10]

Data Mining adalah proses pengolahan data yang digunakan untuk menemukan pola yang tersembunyi dari data tersebut. *Pattern Recognition* merupakan nama lain dari *data mining*. Karena *Data Mining* dapat mengolah data dengan ukuran yang sangat besar, maka *Data Mining* ini memiliki peranan yang sangat besar di dalam berbagai macam bidang, baik industri, teknologi, ilmu pengetahuan bahkan keuangan. Secara umum Metode-Metode yang dibahas dalam *Data Mining* yaitu *Classification*, *Regresi*, *Clustering*, *Market Basket Analisis* dan seleksi variabel[11]. Prinsip teorema bayes dengan Metode *Naive Bayes* ini memiliki variabel yang saling berhubungan. Menghitung *probabilitas* sebuah kejadian

merupakan contoh pendekatan dengan menggunakan teorema *bayes* [12].



Gambar 1. Tahapan *Data Mining*[12]

CRISP-DM dikembangkan oleh analisis dari beberapa industri yang dilakukan pada tahun 1996. CRISP-DM menyediakan standar dari proses *Data Mining* sebagai strategi pemecahan masalah secara umum dari bisnis untuk sebuah penelitian [13]. CRISP-DM merupakan salah satu metode *Data Mining* yang dikembangkan melalui upaya konsorsium yang awalnya disusun oleh Daimler Chrysler, SPSS, dan NCR. Berikut tahapan yang ada dalam proses CRISP-DM [14].

Naive bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Tujuan klasifikasi adalah untuk secara akurat memprediksi kelas target untuk setiap kasus dalam data.[15] Algoritma Naive Bayes didasarkan pada kondisional probabilitas[16]. Sedangkan dasar dari *Naive Bayes* dinyatakan dalam persamaan berikut [17] :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis *H* berdasar kondisi *X* (posteriori probabilitas)

P(H) : Probabilitas hipotesis *H* (prior probabilitas)

P(X|H) : Probabilitas *X* berdasarkan kondisi pada hipotesis *H*

P(X) : Probabilitas *X*

Metode Pengumpulan data yang akan digunakan dalam penelitian ini adalah sebagai berikut:

a. Wawancara

Pengumpulan data yang dilakukan berupa wawancara dengan Pendamping PKH, Koordinator Kecamatan Cimanggu, Koordinator Kabupaten Cilacap mengenai PKH, diantaranya wawancara mengenai bagaimana proses penyaluran PKH, syarat penerima PKH, dan bagaimana melakukan proses survei untuk validasi penerima PKH. Data yang di dapat berupa data Warga miskin yang berada di wilayah kecamatan Cimanggu. Data tersebut didapat dari data terpadu penanganan fakir miskin yang ada di dinas sosial kabupaten cilacap.

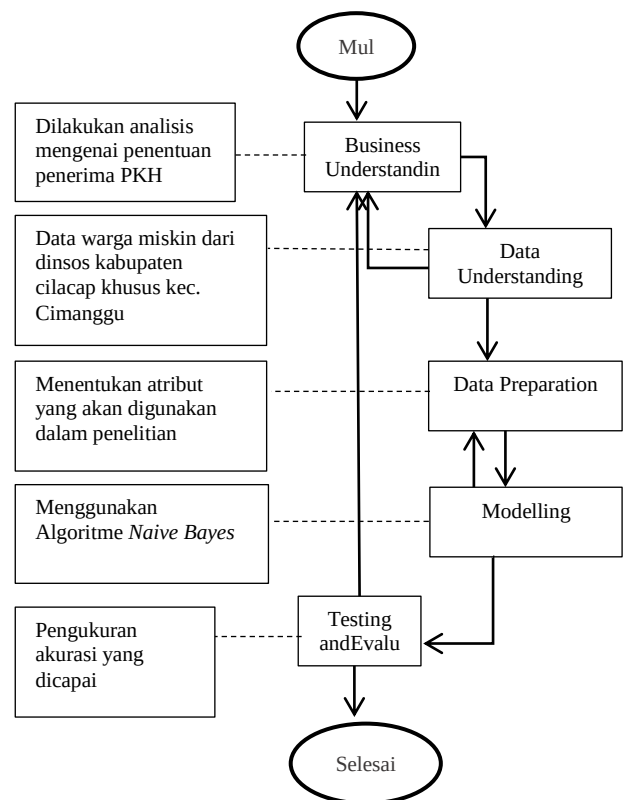
b. Dokumentasi

Dokumentasi penelitian yang didapat antara lain data warga miskin yang ada di wilayah kecamatan cimanggu, syarat menerima PKH, dan UU yang mengatur tentang penerima PKH.

c. Studi Pustaka

Metode penelitian yang digunakan untuk studi pustaka yaitu menggunakan buku, jurnal, artikel, serta kumpulan materi yang berhubungan dengan PKH, *Naive Bayes Classification*, dan *data mining*.

1.1.1. DESAIN PENELITIAN



Gambar 2. Model Penelitian

Desain penelitian yang digunakan pada penelitian ini menggunakan model CRISP-DM (*Cross Industry Standard Process-Data Mining*) yang ada di Gambar 2 memiliki langkah-langkah, yaitu [18] :

a. Bussiness Understanding

PKH merupakan salah satu program bantuan sosial bersyarat dari pemerintah yang ditujukan kepada warga miskin salah satunya warga miskin yang ada di wilayah kecamatan Cimanggu. Jumlah penduduk di wilayah kecamatan Cimanggu sebanyak 97.122 jiwa, dengan jumlah warga miskin sebanyak 14.380 jiwa per Januari 2019. Pada penelitian ini akan dilakukan analisis mengenai penentuan warga miskin yang berhak menerima bantuan PKH supaya bantuan tersebut tepat sasaran.

b. Data Understanding

Data yang didapat pada penelitian ini diperoleh dari data terpadu penanganan faskir miskin Dinas Sosial Kabupaten Cilacap, khususnya wilayah kecamatan cimanggu per Januari 2019. Data tersebut berisikan atribut antara lain : ID BDT, No. Peserta BDT, Kode penduduk, Kode kabupaten, Kode kecamatan, Kode Desa, Alamat, No. Peserta PKH, No. Peserta KKS 2016, No. Peserta PBI, No. Peserta KIP, No. Peserta KIS, Nama, Jumlah ART (Anggota Rumah Tangga), Status keluarga, Status bangunan, status lahan, luas lantai, kondisi dinding, kondisi atap, jumlah kamar, sumber air minum, cara memperoleh air minum, sumber penerangan, fasilitas BAB, jenis kloset, buang tinja, tabung gas, lemari es, ada ac, ada pemanas, ada telepon, ada laptop, ada sepeda, ada sepeda motor, ada mobil, ada perahu, ada kapal, aset tak bergerak, rumah lain, jumlah sapi, jumlah kerbau, jumlah kambing, jumlah babi, usaha, status PKH.

c. Data Preparation

Dari 48 atribut yang ada di *Data Understanding* hanya akan digunakan 13 atribut untuk data penelitian yaitu : No. peserta BDT, Jumlah ART (Anggota Rumah Tangga), Status bangunan, status lahan, jenis lantai, jenis dinding, jenis atap, jumlah kamar, sumber air minum, sumber penerangan, jenis kloset, aset tak bergerak, status PKH.

d. Modelling

Metode *Naive Bayes Classification* merupakan Metode yang akan digunakan pada penelitian ini. Hal ini dikarenakan

metode *Naive Bayes* hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian, dan hal tersebut merupakan keuntungan dari penggunaan Metode *Naive Bayes*. Kebanyakan situasi dalam dunia nyata *Naive Bayes* sering bekerja jauh lebih baik dari pada yang diharapkan.

e. Validasi dan Evaluasi

Pada tahapan ini dilakukan pengukuran tingkat keakuratan hasil dan validasi yang dicapai.

Hasil Dan Pembahasan Bussiness Understanding Phase (Fase Pemahaman Bisnis)

Fase pemahaman bisnis ini berfokus pada tujuan untuk menentukan proyek apa yang sedang dilakukan dan tehnik yang digunakan dalam *Data Mining* yang dilakukan secara detail. Kegiatan pada fase ini antara lain menentukan kebutuhan secara detail dalam lingkup bisnis, menentukan tujuan dan sasaran bisnis, dan menyiapkan strategi untuk mencapai tujuan.

Strategi awal yang digunakan untuk mencapai tujuan adalah dengan melakukan permintaan data Warga Miskin yang ada di Kecamatan Cimanggu dari Dinas Sosial Kabupaten Cilacap. Data yang diberikan dari Dinas Sosial tentang warga miskin per Januari 2019. Tujuan dalam penelitian ini adalah untuk mengklasifikasikan warga yang tidak mampu/miskin yang berhak untuk mendapatkan PKH di kecamatan Cimanggu dengan menggunakan metode *Naive Bayes*. Selain itu juga untuk menentukan atribut apa yang akan digunakan untuk mengklasifikasikan warga tidak mampu/miskin yang berhak menerima/mendapatkan PKH.

Kebutuhan, asumsi dan batasan pada penelitian ini antara lain :

- Data warga miskin/tidak mampu dari Dinas Sosial Kabupaten Cilacap khususnya wilayah Kecamatan Cimanggu.
- Metode yang digunakan dalam penelitian ini adalah Metode *Naive Bayes*.
- Hasil penelitian ini dapat digunakan sebagai acuan bagi pemerintah daerah terutama Dinas Sosial dalam menanggulangi permasalahan penerima

bantuan PKH yang sesuai dengan aturan Pemerintah yang berlaku.

- d. Data yang akan digunakan dalam penelitian ini sudah mendapatkan izin dari Dinas Sosial Kabupaten Cilacap

Data Understanding Phase (Fase Pemahaman Data)

Data yang digunakan dalam penelitian ini adalah Data Warga Miskin Kecamatan Cimanggu Per Januari 2019 yang di dapat dari Dinas Sosial Kabupaten Cilacap. Data yang didapat berjumlah 14.832 data. Selain itu juga dilakukan wawancara dengan pendamping PKH, Koordinator kabupaten bagian PKH, dan supervisor Dinas Sosial terkait penjelasan data yang telah diberikan dan bagaimana proses PKH ini berjalan dari awal menentukan warga miskin yang berhak, sampai dengan proses penerimaan bantuan kepada Keluarga Penerima Manfaat PKH. Proses-proses tersebut sesuai dengan permensos no.1 tahun 2018.

TABEL I.
PENJELASAN ATRIBUT DATA

No	Atribut	Keterangan
1	No. Peserta BDT	Berisi No. Basis Data Terpadu yang dimiliki oleh setiap warga di Indonesia.
2	No. Peserta PKH	Berisi No. Peserta PKH
3	No. Peserta KKS 2016	Berisi No. Peserta Kartu Keluarga Sejahtera
4	No. Peserta PBI	Berisi No. Peserta Penerima Bantuan Iuran
5	Jumlah ART	Berisi Jumlah Anggota Rumah Tangga dalam 1 KK
6	Status Bangunan	Berisi Status kepemilikan Bangunan yang ditinggali
7	Status Lahan	Berisi Status kepemilikan Lahan dari bangunan yang ditinggali
8	Jenis Lantai	Berisi jenis lantai yang pada bangunan yang ditinggali
9	Jenis Dinding	Berisi jenis dinding dari bangunan yang ditinggali
10	Jenis Atap	Berisi jenis Atap dari bangunan yang ditinggali
11	Jumlah kamar	Berisi jumlah kamar yang ada dari bangunan yang ditinggali
12	Sumber air minum	Berisi tentang informasi mengenai sumber air

No	Atribut	Keterangan
		minum yang biasa digunakan oleh Warga
13	Cara Peroleh Air minum	Berisi informasi mengenai cara memperoleh air minum yang digunakan oleh warga
14	Sumber penerangan	Berisi informasi mengenai sumber penerangan yang digunakan pada bangunan yang ditinggali
15	Fasbab	Berisi informasi mengenai fasilitas buang air besar
16	Jenis Kloset	Berisi informasi mengenai jenis kloset yang digunakan
17	Buang tinja	Berisi informasi mengenai pembuangan tinja
18	Ada tabung gas	Berisi informasi kepemilikan tabung gas
19	Ada lemari es	Berisi informasi kepemilikan lemari es
20	Ada ac	Berisi informasi kepemilikan ac
21	Ada pemanas	Berisi informasi kepemilikan pemanas
22	Ada telepon	Berisi informasi kepemilikan telepon
23	Ada tv	Berisi informasi kepemilikan televisi
24	Ada emas	Berisi informasi kepemilikan emas
25	Ada laptop	Berisi informasi kepemilikan laptop
26	Ada sepeda	Berisi informasi kepemilikan sepeda
27	Ada motor	Berisi informasi kepemilikan motor
28	Ada mobil	Berisi informasi kepemilikan mobil
29	Ada perahu	Berisi informasi kepemilikan perahu
30	Ada motor tempel	Berisi informasi kepemilikan motor tempel
31	Ada perahu motor	Berisi informasi kepemilikan perahu motor
32	Ada kapal	Berisi informasi kepemilikan kapal
33	Aset tak bergerak	Berisi informasi kepemilikan aset tak bergerak termasuk lahan dan rumah lain
34	Jumlah sapi	Berisi informasi jumlah sapi yang dimiliki
35	Jumlah kerbau	Berisi informasi jumlah kerbau yang dimiliki

No	Atribut	Keterangan
36	Jumlah kuda	Berisi informasi jumlah kuda yang dimiliki
37	Jumlah babi	Berisi informasi jumlah babi yang dimiliki
38	Jumlah kambing	Berisi informasi jumlah kambing yang dimiliki
39	Status usaha	Berisi informasi status usaha yang dimiliki
40	Status kks	Berisi informasi status Kartu Keluarga Sejahtera
41	Status kip	Berisi informasi status Kartu Indonesia Pintar
42	Status kis	Berisi informasi status Kartu Indonesia Sehat
43	Status bpjs mandiri	Berisi informasi status BPJS Mandiri
44	Status jamsostek	Berisi informasi status jamsostek
45	Status asuransi	Berisi informasi status asuransi
46	Status PKH	Berisi informasi status PKH
47	Status rastra	Berisi informasi status penerimaan Beras Rastra
48	Status kur	Berisi informasi Kredit Usaha Rakyat

Verifikasi Kualitas Data

Pada tahapan ini dilakukan pemeriksaan terhadap kualitas data yang akan digunakan diantaranya pemeriksaan kualitas data, data yang kosong dan kemiripan data. Hasil dari tahapan ini digunakan sebagai dasar untuk melakukan proses verifikasi terhadap kualitas data yang akan digunakan. Dari tahap tersebut diperoleh ada beberapa pada data warga miskin memiliki beberapa *missing value* dan kesamaan data/*Distinct* yang dijelaskan pada tabel 2.

TABEL II.
DATA MISSING VALUE DAN DISTINCT

Atribut	Deskripsi	Jumlah
Jumlah ART	Missing Value	0
	Distinct	9
Status Bangunan	Missing Value	321
	Distinct	5
Status Lahan	Missing Value	563
	Distinct	4
Jenis Lantai	Missing Value	321
	Distinct	10
Jenis Dinding	Missing Value	321
	Distinct	7
Jenis Atap	Missing Value	2399
	Distinct	9
Jumlah Kamar	Missing Value	479
	Distinct	9
Sumber Air Minum	Missing Value	0
	Distinct	13

Atribut	Deskripsi	Jumlah
Cara Peroleh Air Minum	Missing Value	321
	Distinct	3
Sumber Penerangan	Missing Value	321
	Distinct	3
Fasbab	Missing Value	321
	Distinct	4
Kloset	Missing Value	3395
	Distinct	4
Buang Tinja	Missing Value	321
	Distinct	6
Ada tabung gas	Missing Value	321
	Distinct	2
Ada lemari es	Missing Value	494
	Distinct	2
Ada ac	Missing Value	321
	Distinct	2
Ada pemanas	Missing Value	494
	Distinct	2
Ada Telepon	Missing Value	321
	Distinct	2
Ada TV	Missing Value	494
	Distinct	2
Ada Emas	Missing Value	321
	Distinct	2
Ada Laptop	Missing Value	494
	Distinct	2
Ada Sepeda	Missing Value	321
	Distinct	2
Ada Motor	Missing Value	494
	Distinct	2
Ada Mobil	Missing Value	321
	Distinct	2
Ada Perahu	Missing Value	494
	Distinct	2
Ada Motor Tempel	Missing Value	321
	Distinct	2
Ada Perahu Motor	Missing Value	494
	Distinct	2
Ada Kapal	Missing Value	321
	Distinct	2
Aset Tak Bergerak	Missing Value	322
	Distinct	4
Jumlah Sapi	Missing Value	0
	Distinct	6
Jumlah Kerbau	Missing Value	0
	Distinct	5
Jumlah Kuda	Missing Value	0
	Distinct	2
Jumlah Babi	Missing Value	0
	Distinct	5
Jumlah Kambing	Missing Value	0
	Distinct	14
Status PKH	Missing Value	321
	Distinct	2

Data Preparation Phase (Fase Pengolahan Data)

Pada fase pengolahan data disiapkan kumpulan data yang dibutuhkan yang akan

digunakan pada fase berikutnya. Fase ini merupakan pekerjaan yang sangat berat dan perlu dilakukan secara intensif. Siapkan variabel yang akan ditransformasikan pada database baru dan lakukan perubahan variabel jika itu sangat dibutuhkan. *Data Preparation Phase* meliputi seluruh aktifitas yang dilakukan untuk membangun dataset yang akan digunakan sebagai masukan pada aplikasi pemodelan dari data mentah. Persiapan data yang diperlukan biasanya dilakukan berulang kali untuk memastikan kualitas data.

Pembersihan Data

Pada tahapan pembersihan data dilakukan proses pembersihan data. Jumlah data yang ada sejumlah 14.382 data. Berdasarkan hasil verifikasi kualitas data yang telah dilakukan diperoleh hasil yaitu :

- Missing Value* banyak ditemukan pada data warga miskin. Sehingga akan dilakukan penghapusan data terhadap data yang kosong. Data akan dihapus jika data pada atribut Status PKH kosong. Data yang dihapus pada atribut status PKH sejumlah 321 data. Data menjadi berjumlah 14.061 data
- Missing Value* pada atribut status lahan berjumlah 242. Data yang kosong tersebut akan dihapus. Sehingga jumlah data menjadi 13.819 data.
- Missing Value* pada atribut jenis atap yang berjumlah 2013 data. Data tersebut akan dihapus. Sehingga jumlah data menjadi 11.806 data.
- Missing Value* pada atribut Kloset yang berjumlah 2179 data. Data tersebut akan dihapus, sehingga jumlah data menjadi 9.627 data.
- Kemudian untuk *Missing Value* pada atribut ada lemari es yang berjumlah 172 data juga akan dilakukan penghapusan. Sehingga data yang akan digunakan berjumlah 9.455 data.

Pemilihan Data

Penelitian ini hanya menggunakan 13 atribut dari 47 atribut yang ada, yaitu No.peserta BDT, Jumlah ART, Status bangunan, status lahan, jenis lantai, jenis dinding, jenis atap, jumlah kamar, sumber air minum, sumber penerangan, Jenis Kloset, aset tak bergerak, dan status PKH. Atribut ini di pilih karena berdasarkan Kepmensos No. 146 tahun 2013 Tentang Penetapan Kriteria dan Pendataan Fakir Miskin dan Orang Tidak mampu.

TABEL III.
ATRIBUT DATA PENELITIAN

Atribut	Deskripsi
No. Peserta BDT	Berisi informasi mengenai nomor Basis Data Terpadu (BDT).
Jumlah ART	Berisi informasi mengenai Jumlah Anggota Rumah Tangga dalam 1 KK
Status Lahan	Berisi Informasi mengenai kepemilikan lahan dari bangunan yang ditinggali
Jenis lantai	Berisi informasi mengenai jenis lantai dari bangunan yang ditinggali
Jenis Dinding	Berisi jenis dinding dari bangunan yang ada
Jenis Atap	Berisi jenis atap dari bangunan yang ada
Sumber Air Minum	Berisi informasi mengenai sumber air minum yang digunakan oleh warga
Cara Peroleh Air MInum	Berisi informasi mengenai cara memperoleh air minum warga.
Sumber Penerangan	Berisi Informasi mengenai sumber penerangan yang digunakan pada bangunan yang ditinggali
Jenis Kloset	Berisi Informasi jenis kloset yang digunakan
Ada lemari es	Berisi informasi kepemilikan lemari es
Ada telepon	Berisi informasi kepemilikan telepon
Ada laptop	Berisi informasi kepemilikan laptop
Ada perahu motor	Berisi informasi kepemilikan perahu motor
Ada kapal	Berisi informasi kepemilikan kapal
Aset Tak bergerak	Berisi informasi kepemilikan aset tak bergerak
Status PKH	Berisi informasi status penerimaan Bantuan PKH

Modelling Phase (Fase Pemodelan)

Metode *Naive Bayes Classification* merupakan metode yang akan digunakan pada penelitian ini. Sedangkan pengukuran tingkat akurasi menggunakan *Framework* Weka ver. 3.9.3. pada tahapan ini dilakukan pemilihan teknik pemodelan, pembuatan model, serta pengujian model. Teknik

pemodelan yang akan digunakan untuk melakukan klasifikasi pada warga tidak mampu/miskin di kecamatan cimanggu adalah *Naive Bayes*. Sedangkan aplikasi yang digunakan yaitu aplikasi weka ver 3.9.3.

Tahap selanjutnya yang perlu dilakukan adalah melakukan klasifikasi data dengan menggunakan metode *naive bayes* pada aplikasi weka. Tahapan ini mencoba melakukan percobaan untuk membandingkan tingkat akurasi dengan *percentage split* yang berbeda dengan menggunakan metode *Naive Bayes*.

TABEL IV.

PERCOBAAN HASIL AKURASI METODE NAIVE BAYES

Percentage Split %	Akurasi %	Presisi	Recall
10%	82,4304	0,844	0,969
20%	83,0249	0,843	0,979
30%	83,3182	0,844	0,981
40%	83,8710	0,848	0,983
50%	84,0914	0,850	0,984
60%	84,4791	0,850	0,980
70%	84,1678	0,849	0,980
80%	82,9191	0,830	0,980
90%	83,1746	0,830	0,992

Tabel 4 menunjukkan bahwa hasil pengujian dengan menggunakan metode *Naive Bayes* mampu melakukan proses klasifikasi dengan tingkat akurasi tertinggi

dengan percentage split sebesar 60% dengan tingkat akurasi sebesar 84,4791%, presisi 0,85 dan *recall* 0,98. Sedangkan tingkat akurasi terendah didapat dengan *percentage split* 10% dengan tingkat akurasi sebesar 82,4304 % dan nilai *presisi* sebesar 0,84 sedangkan nilai *recall* 0,96. Kinerja klasifikasi dari sebuah metode dapat dianggap baik jika nilai *presisi* dan nilai *recall* memiliki nilai yang tinggi, tetapi selain itu juga perlu diperhatikan tingkat akurasinya. Klasifikasi dengan menggunakan metode *Naive Bayes* dengan *percentage split* 60%.

Validasi dan Evaluasi

Pada tahapan ini dilakukan pengukuran tingkat keakuratan hasil dan validasi yang dicapai oleh model CRISP-DM dengan menggunakan beberapa tehnik yang terdapat dalam *Framework Weka* ver. 3.9.3. Pengujian ini dilakukan dengan *percentage split* sebesar 60%. *Persentase* tersebut didapat dari hasil percobaan akurasi yang terdapat pada tabel 4. Pada percobaan yang dilakukan, dihasilkan bahwa percobaan dengan menggunakan *percentage split* sebesar 60% menghasilkan tingkat akurasi paling tinggi.

```

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===
Correctly Classified Instances      3195      84.4791 %
Incorrectly Classified Instances    587      15.5209 %
Kappa statistic                    0.1641
Mean absolute error                 0.2327
Root mean squared error             0.3422
Relative absolute error             85.7548 %
Root relative squared error         92.8625 %
Total Number of Instances          3782

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,983	0,871	0,854	0,983	0,914	0,225	0,760	0,935	Tidak
	0,129	0,017	0,598	0,129	0,212	0,225	0,760	0,396	Ya
Weighted Avg.	0,845	0,733	0,812	0,845	0,800	0,225	0,760	0,848	

```

=== Confusion Matrix ===
a    b    <-- classified as
3116  53 |    a = Tidak
 534 79 |    b = Ya

```

Gambar 3. hasil pengujian *Naive Bayes* dengan *percentage split* 60%

Berdasarkan *Confusion Matrix* yang terdapat dalam gambar 3, kesalahan/error tersebut berasal dari 53 data dengan *Class* tidak yang salah diklasifikasikan kedalam *Class* Ya, dan 534 data dengan *Class* ya yang salah diklasifikasikan kedalam *Class* tidak.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut :

- Data warga miskin yang didapat dari Dinas Sosial Kabupaten Cilacap dengan

percobaan menggunakan *percentage split* sebesar 60% dapat disimpulkan bahwa dari jumlah 3782 data warga miskin, yang berhak menerima PKH sebanyak 613 warga dan yang tidak berhak untuk menerima PKH sebanyak 3169 warga.

- Tingkat akurasi data setelah dilakukan evaluasi dengan menggunakan *Confusion Matrix* adalah sebesar 84,2411% yang masuk dalam kategori *Good Classification*.

- c. Atribut yang digunakan dalam penelitian ini berjumlah 17 yaitu No Peserta BDT, Jumlah ART, status lahan, jenis lantai, jenis dinding, jenis atap, sumber air minum, Cara peroleh air minum, sumber penerangan, jenis kloset, ada lemari es, ada telepon, ada laptop, ada perahu motor, ada kapal, aset tak bergerak, dan status PKH
- d. Atribut yang paling berpengaruh dalam menentukan warga tersebut berhak menerima PKH atau tidak, dipengaruhi oleh kepemilikan kapal, laptop, perahu motor, telepon, lemari es, cara memperoleh air minum, sumber penerangan, jenis atap, status lahan, jenis kloset, jenis lantai, aset tak bergerak dan jumlah ART, Karena jika atribut tersebut dihilangkan, tingkat akurasiya berkurang.
- e. Dari hasil klasifikasi di dapatkan pola bahwa warga yang memiliki Jumlah ART (Anggota Rumah Tangga) di atas 3 dan tidak memiliki lemari es, telepon, laptop, perahu motor, kapal dan aset tak bergerak cenderung mendapatkan PKH.
- f. Status kepemilikan lahan dengan status milik sendiri, lebih dominan tidak mendapatkan PKH.

REFERENSI

- [1] S. Heni and A. Irham Gufroni, "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia," *J. Nas. Teknol. dan Sist. Inf.*, vol. 03, no. 02, pp. 299–305, 2017.
- [2] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
- [3] TNP2K, "Program Bantuan Pemerintah Untuk Individu, Keluarga dna Kelompok Tidak Mampu menuju bantuan sosial Terintegrasi ringkasan eksekutif," in *TNP2K*, Jakarta: TNP2K, 2018.
- [4] C. Fadlan, S. Ningsih, and A. P. Windarto, "Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *J. Tek. Inform. Musirawas*, vol. 3, no. 1, p. 1, 2018.
- [5] mohamad jajuli nurul rohmawati, sofi defiyanti, "Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa," *Jitter 2015*, vol. I, no. 2, pp. 62–68, 2015.
- [6] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, p. 160, 2018.
- [7] M. Jamaris, "Implementasi Metode Rough Set Untuk Menentukan Kelayakan Bantuan Dana Hibah Fasilitas Rumah Ibadah," *INOVTEK Polbeng - Seri Inform.*, vol. 2, no. 2, p. 161, 2017.
- [8] D. Iskandar and Y. K. Suprpto, "Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4 . 5 dan Naïve Bayes Clasifier," *JAVA J. Electr. Electron. Eng.*, vol. 11, no. 1, pp. 14–17, 2013.
- [9] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [10] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "Implementasi Algoritma C5 . 0 Dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus : Kelurahan Mesjid Kecamatan Medan Kota).," *Inf. dan Teknol. Ilm.*, vol. 13, pp. 192–198, 2018.
- [11] R. T. Vlandari, "Pengelompokan Tingkat Keamanan Wilayah Jawa Tengah Berdasarkan Indeks Kejahatan Dan Jumlah Pos Keamanan Dengan Metode Klastering K-Means," *J. Ilm. SINUS*, vol. Vol 14, No, no. ISSN :1693-1173, pp. 59–72, 2016.
- [12] Y. Lukito and A. R. Chrismanto, "Perbandingan Metode-Metode Klasifikasi untuk Indoor Positioning System," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 123–131, 2015.
- [13] D. Listriani, A. H. Setyaningrum, and F. Eka, "Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro)," *J. Tek. Inform.*, vol. 9, no. 2, pp. 120–127, 2018.
- [14] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*, vol. 9780470908. 2014.
- [15] Irmawati, "Data Mining Untuk Penentuan Model Tingkat Kesuksesan Kelulusan Murid Sma Pada Perguruan Tinggi Negeri : Data Mining For Determination Of High School Student Graduation Model At State University ; Case Study In Iain Bone," vol. 3, no. 2, pp. 113–118, 2020.
- [16] Sumantri, R. Bagus Bambang, and Ema Utami. "Determination of status of family stage prosperous of Sidareja district using data mining techniques." *International Journal of Intelligent Systems and Applications* 10.10 (2018): 1.
- [17] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *J. Inform. Ahmad Dahlan*, vol. 8, no. 1, p. 102632, 2014.
- [18] Muhardi, "Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen , Motivasi , Kedisiplinan , Ekonomi ,dan Hasil Belajar," vol. 2, no. 2, pp. 41–44, 2016.