

Klasifikasi Pendapatan Pedagang Kaki Lima Dan Pelaku Usaha *Online* Akibat Dampak *Covid-19* Menggunakan Metode *Naive Bayes*

L. Yuwono¹, M.E. Fadillah², M. Indrayani³, W. Maesarah⁴, A. Ramadhan⁵, S.F. Panjaitan⁶

Abstrak— Pandemi *Corona Virus* atau dikenal juga dengan sebutan *Covid-19* yang terjadi di seluruh dunia dan tidak terkecuali Indonesia mengakibatkan ketidakstabilan perekonomian terutama pendapatan pedagang kaki lima dan pelaku usaha *online*. Penelitian melakukan klasifikasi berdasarkan data pendapatan menurun dari pelaku usaha yang diperoleh dari pengisian *Kuesioner* dengan teknik *data mining*. Atribut yang akan digunakan dalam melakukan klasifikasi pendapatan adalah Pendidikan, Jenis Usaha, Penghasilan, Status (Menikah/Lajang). Metode yang akan digunakan adalah metode *Naive Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam *data mining*. Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa pendapatan pedagang kaki lima dan pelaku usaha *online* menurun sebesar 89%. Kemudian, hasil pengujian berdasarkan *confusion matrix* dengan teknik *split validasi*, penggunaan metode klasifikasi *naive bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 75% atau termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 100% dan *Recall* sebesar 66.7%.

Kata Kunci— Data Mining, Klasifikasi, *Naive Bayes*, Tingkat Pendapatan menurun.

Abstract — The Pandemi *Corona Virus*, also known as the *Covid-19*, takes place worldwide and is no exception to Indonesia resulting in economic instability, especially the income business of street and online businesses. The research is based on declining revenue data from business actors obtained from the filling of questionnaires with data mining techniques. The attributes to be used in the classification of income are education, type of business, income, Status (Married/Single). The method to be used is the *Naive Bayes Classifier* method, which is one of the classifying techniques in data mining. Based on the research made the conclusion that the income of street vendors and online business decreased by 89%. Then, the results of testing based on *confusion matrix* with the technique of *split validation*, use of classification method *naive Bayes* to the dataset that has been taken on the research object obtained an accuracy rate of 75% or belongs to the category *Good*. While the *Precision* value is 100% and the *Recall* is 66.7%.

Kata Kunci— Data Mining, Klasifikasi, *Naive Bayes*, Tingkat Pendapatan menurun.

I. PENDAHULUAN

Awal tahun 2020 dunia dikejutkan dengan adanya Asebuah wabah yang meresahkan seluruh dunia, tak terkecuali Indonesia. Permasalahan tersebut banyak membuat warga dunia dan juga Indonesia resah. Karena wabah *Corona Virus* atau *Covid-19* ini banyak Negara yang menerapkan *lockdown* atau isolasi untuk membatasi penyebaran *Covid-19*, dengan kondisi ini aktivitas warga menjadi terbatas terutama di kota-kota besar seperti Jakarta yang berimbas kepada perekonomian pedagang kaki lima dan pelaku usaha *online* yang mengalami penurunan pendapatan akibat dampak dari *Covid-19*. Upaya penanganan saat ini terkait permasalahan tersebut dengan adanya fakta bahwa pendapatan pedagang kaki lima maupun pelaku usaha *online* warga Jakarta secara merata

menurun yang disebabkan aktivitas warga yang sangat terbatas akibat wabah *Covid-19* [1]. Data klasifikasi yang akan digunakan adalah data pedagang kaki lima dan pelaku usaha *online* berdasarkan Status, Pendidikan, jenis usaha, Penghasilan dan data pendapatan pedagang kaki lima maupun pelaku usaha *online* yang mengalami penurunan dan stabil. Data penelitian diperoleh dari pengisian kuesioner yang dilakukan di *social media* secara *random*. Metode yang akan digunakan adalah metode *Naive Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam *data mining* [2]. Rumusan masalah penelitian ini adalah menentukan cara merekayasa sistem untuk klasifikasi pendapatan Pedagang kaki lima dan Pelaku usaha *Online* yang mengalami pendapatan menurun selama pandemi *Covid-19*, menggunakan *Algoritma Naive Bayes* dan hasil penerapan *Algoritma Naive Bayes* untuk di klasifikasi pendapatan menurun yang diperoleh dari pengisian kuesioner di *social media* secara *random*. Adapun tujuan dari penelitian ini adalah klasifikasi pendapatan pedagang kaki lima dan pelaku usaha *online*, menggunakan *Algoritma Naive Bayes* dan menerapkan *Algoritma Naive Bayes* untuk klasifikasi pendapatan pedagang kaki lima dan pelaku usaha *online* yang diperoleh dari pengisian *kuesioner* di *social media*.

L. Yuwono, Mahasiswa Program Studi Teknik Industri, Universitas Indraprasta PGRI, Jakarta (ludiroyuwono@gmail.com).

M.E. Fadillah, Mahasiswa Program Studi Teknik Industri, Universitas Indraprasta PGRI, Jakarta (muhamadegifadillah10@gmail.com).

M. Indrayani, Mahasiswa Program Studi Teknik Industri, Universitas Indraprasta PGRI, Jakarta (muhammadindra606@gmail.com).

W. Maesarah, Mahasiswa Program Studi Teknik Industri, Universitas Indraprasta PGRI, Jakarta (widjanmaesarah@gmail.com).

A. Ramadhan, Mahasiswa Program Studi Teknik Industri, Universitas Indraprasta PGRI, Jakarta (Amillramadhan12@gmail.com).

S.F. Panjaitan, Mahasiswa Program Studi Teknik Industri, Universitas Indraprasta PGRI, Jakarta (Friscasaswita96@gmail.com).

II. METODE DAN PROSEDUR

A. Klasifikasi

Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (SVM)* [3].

B. Algoritma Naive Bayes

Naïve Bayes classifiers adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [4]. *Naïve Bayes classifiers* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. *Naïve Bayes classifiers* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode *Bayes* merupakan pendekatan *statistic* untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema *Bayes*, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam *data mining* [5]. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan

- 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

Pengujian terhadap metode klasifikasi *naïve bayes* yang digunakan, dilakukan dengan menggunakan teknik *split validation* dengan *confusion matrix*.

1. Split Validation

Split Validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai *data training* dan sebagian lainnya sebagai *data testing*, dengan menggunakan *Split Validation* akan dilakukan percobaan *training* berdasarkan *split ratio* yang telah ditentukan sebelumnya, untuk kemudian sisa dari *split ratio data training* akan dianggap sebagai *data testing*. *Data training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan *data testing* adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran [6].



Gambar 1. Ilustrasi Split Validation

2. Confusion Matrix

Confusion matrix juga sering disebut *error matrix*. Pada dasarnya *confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui [7].

Berikut adalah beberapa manfaat dari *confusion matrix*:

1. Menunjukkan bagaimana model ketika membuat prediksi.
2. Tidak hanya memberi informasi tentang kesalahan yang dibuat oleh model tetapi juga jenis kesalahan yang dibuat.
3. Setiap kolom dari *confusion matrix* mempresentasikan *instance* dari kelas prediksi.
4. Setiap baris dari *confusion matrix* mewakili *instance* dari kelas aktual.

Tabel 1
Confusion Matrix 2 Kelas

Classification	Predicted Class	
	Class=Yes	Class=No
Actual=Yes	(True Positive-TP)	(False Negative-FN)
Actual= No	(False Positive-FP)	(True Negative-TN)

- TP (*True Positive*) → Merupakan data positif yang diprediksi benar.
- FP (*False Positive*) → Merupakan data negatif namun diprediksi sebagai data positif.
- FN (*False Negative*) → Merupakan data positif namun diprediksi sebagai data negatif.
- TN (*True Negative*) → Merupakan data negatif yang diprediksi benar.

3. Accuracy

Accuracy menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar. Maka, *Accuracy* merupakan prediksi benar (positif dan negatif) dengan keseluruhan data. *Accuracy* merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya). Nilai *accuracy* dapat diperoleh dengan persamaan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

4. Precision

Precision menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Maka, *precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari semua kelas positif yang telah di prediksi dengan benar, berapa banyak data yang benar-benar positif. Nilai *precision* dapat diperoleh dengan persamaan.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

5. Recall

Recall menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Maka, *recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Nilai *recall* dapat diperoleh dengan persamaan.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

C. Prosedur penelitian

1. Identifikasi Masalah

Permasalahan yang diteliti dalam penelitian ini adalah dampak dari Wabah *Covid-19* terhadap pendapatan Pelaku Usaha. Penelitian ini dilakukan untuk mengklasifikasi pendapatan pelaku usaha menurun/stabil setelah pandemi *Covid-19* terjadi.

2. Pengumpulan Data

Pengumpulan data dilakukan secara online melalui *Google Form* dengan cara mengisi *Kuesioner*. Pengumpulan data diperoleh lebih dari 30 orang pelaku usaha. Proses Pengumpulan data dilakukan dengan cara membagikan *link kuisisioner* kepada pelaku usaha, dikarenakan situasi dan kondisi saat ini tidak memungkinkan untuk secara langsung melakukan wawancara kepada pelaku usaha.

Tabel 2
Klasifikasi Pengumpulan Data

No	Umur	Status	Pendidikan	Jenis Usaha	Penghasilan	Pendapatan Setelah Pandemi Covid-19
1	23	Menikah	Perguruan Tinggi	Obat Asam Urat	Sedang	Menurun
2	25	Menikah	SMA/SMK	Makanan	Tinggi	Menurun
3	24	Lajang	SMA/SMK	Makanan	Rendah	Menurun
4	19	Lajang	Perguruan Tinggi	Sepatu	Tinggi	Menurun
5	24	Menikah	SMA/SMK	Makanan	Tinggi	Menurun
6	21	Lajang	Perguruan Tinggi	Sepatu	Rendah	Menurun
7	35	Menikah	SMA/SMK	Makanan	Sedang	Menurun
8	23	Lajang	SMA/SMK	Wedding Accessories	Rendah	Menurun
9	18	Lajang	SMP	Makanan	Tinggi	Menurun
10	23	Lajang	SMA/SMK	Makanan	Sedang	Menurun
11	18	Lajang	SMA/SMK	Makanan	Rendah	Menurun
12	20	Lajang	SMA/SMK	Sepatu	Tinggi	Menurun
13	20	Lajang	Perguruan Tinggi	Cleaning shoes	Rendah	Menurun
14	21	Lajang	Perguruan Tinggi	Online Shop	Rendah	Stabil
15	24	Lajang	Perguruan Tinggi	Makanan	Rendah	Menurun
16	28	Menikah	SMA/SMK	Makanan	Sedang	Menurun
17	19	Lajang	SMA/SMK	T-shirt	Rendah	Menurun
18	21	Lajang	Perguruan Tinggi	Makanan	Tinggi	Menurun
19	20	Lajang	SMA/SMK	T-shirt	Rendah	Menurun
20	24	Lajang	Perguruan Tinggi	Cosmetic	Tinggi	Menurun
21	25	Menikah	SMA/SMK	Obat	Tinggi	Menurun
22	22	Lajang	SMA/SMK	Online Shop	Rendah	Menurun
23	23	Lajang	Perguruan Tinggi	Makanan	Tinggi	Menurun
24	23	Lajang	SMA/SMK	Fashion	Tinggi	Stabil
25	21	Lajang	Perguruan Tinggi	Cosmetic	Sedang	Menurun
26	24	Lajang	SMA/SMK	Makanan	Sedang	Menurun
27	19	Lajang	SMA/SMK	T-shirt	Sedang	Menurun
28	21	Lajang	Perguruan Tinggi	Parfum	Rendah	Stabil
29	24	Lajang	Perguruan Tinggi	Makanan	Tinggi	Menurun
30	26	Menikah	SMA/SMK	Product Nasa	Sedang	Stabil

31	22	Menikah	SMA/SMK	(Cosmetic) Makanan (Kue)	Rendah	Menurun
32	22	Menikah	SMA/SMK	PKL (Makanan)	Tinggi	Menurun
33	25	Menikah	SMA/SMK	PKL (Makanan)	Tinggi	Menurun
34	25	Lajang	Perguruan Tinggi	Makanan	Rendah	Menurun
35	40	Menikah	SMA/SMK	PKL (Makanan)	Sedang	Menurun
36	38	Menikah	SMA/SMK	Makanan Kasur Inoac	Sedang	Menurun
37	23	Menikah	SMA/SMK	(Property) Makanan (Cake)	Rendah	Menurun
38	33	Menikah	SMA/SMK	Olshop menjual luxury watch	Tinggi	Stabil
39	24	Lajang	Perguruan Tinggi	Property	Tinggi	Menurun
40	22	Lajang	Perguruan Tinggi	Property	Tinggi	Menurun

3. Pengolahan Data

Setelah dilakukan pengumpulan data, langkah selanjutnya adalah pengolahan data set menggunakan teknik *Split Validation* dengan *Confusion Matrix*. Dalam pengujian metode klasifikasi *naive bayes* dengan teknik *split validation* dimana dataset yang disajikan pada tabel klasifikasi pengumpulan data yang berjumlah 40 akan dibagi kedalam dua bagian yaitu 90% (36 record) dari Dataset akan dijadikan sebagai data training atau latih dan 10% (4 record) sisanya akan dijadikan sebagai data testing atau uji. Untuk proses klasifikasi terhadap *data training* menggunakan *attribute/Class* [8] berikut:

Status
Pendidikan
Jenis Usaha
Penghasilan

Untuk Pengujian penggunaan metode klasifikasi *naive bayes* dengan *confusion matrix*, dapat diukur dengan menghitung nilai *accuracy*, *precision* dan *recall* [9].

III. HASIL DAN PEMBAHASAN

A. Manual Penggunaan Metode Klasifikasi Dengan Menggunakan Data Training/Latih [10].

Berdasarkan data training/latih, maka akan dilakukan proses klasifikasi pendapatan menurun atau stabil apabila diberikan data baru berikut:

Status = Menikah

Pendidikan = Perguruan Tinggi

Jenis Usaha = Fashion

Penghasilan = Sedang

maka klasifikasi data Pendapatan Pelaku usaha dapat ditentukan melalui langkah berikut :

Proses *naive bayes*

1. Menghitung jumlah class/label

$$P(\text{Menurun}) = \frac{32}{36} = 0,89$$

$$0,89 \times 100\% = 89\%$$

“Jumlah data pendapatan menurun pada data Training dibagi dengan jumlah keseluruhan data dikali 100%”

$$P(\text{Stabil}) = \frac{4}{36} = 0,11$$

$$0,11 \times 100\% = 11\%$$

“Jumlah data Pendapatan Stabil pada data Training dibagi dengan jumlah keseluruhan data dikali 100%”

2. Menghitung jumlah kasus yang sama dengan *class* yang sama

Probabilitas atribut Status:

$$P(\text{Status} = \text{Lajang} | \text{Pendapatan menurun}) = 20/32 = 0.625$$

$$P(\text{Status} = \text{Lajang} | \text{Pendapatan Stabil}) = 3/4 = 0,75$$

$$P(\text{Status} = \text{Menikah} | \text{Pendapatan menurun}) = 12/32 = 0.375$$

$$P(\text{Status} = \text{Menikah} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

Probabilitas atribut Pendidikan:

$$P(\text{Pendidikan} = \text{SMA/SMK} | \text{Pendapatan menurun}) = 20/32 = 0.625$$

$$P(\text{Pendidikan} = \text{SMA/SMK} | \text{Pendapatan Stabil}) = 2/4 = 0.5$$

$$P(\text{Pendidikan} = \text{Perguruan Tinggi} | \text{Pendapatan Menurun}) = 11/32 = 0.343$$

$$P(\text{Pendidikan} = \text{Perguruan Tinggi} | \text{Pendapatan Stabil}) = 2/4 = 0.5$$

$$P(\text{Pendidikan} = \text{SMP} | \text{Pendapatan Menurun}) = 1/32 = 0.03$$

$$P(\text{Pendidikan} = \text{SMP} | \text{Pendapatan Stabil}) = 0/4 = 0$$

Probabilitas atribut Jenis Usaha:

$$P(\text{Jenis Usaha} = \text{Makanan} | \text{Pendapatan Menurun}) = 19/32 = 0.593$$

$$P(\text{Jenis Usaha} = \text{Makanan} | \text{Pendapatan Stabil}) = 0/4 = 0$$

$$P(\text{Jenis Usaha} = \text{Sepatu} | \text{Pendapatan Menurun}) = 3/32 = 0,09$$

$$P(\text{Jenis Usaha} = \text{Sepatu} | \text{Pendapatan Stabil}) = 0/4 = 0$$

$$P(\text{Jenis Usaha} = \text{T-shirt} | \text{Pendapatan Menurun}) = 2/32 = 0,06$$

$$P(\text{Jenis Usaha} = \text{T-shirt} | \text{Pendapatan Stabil}) = 0/4 = 0$$

$$P(\text{Jenis Usaha} = \text{Obat} | \text{Pendapatan Menurun}) = 2/32 = 0,06$$

$$P(\text{Jenis Usaha} = \text{Obat} | \text{Pendapatan Stabil}) = 0/4 = 0$$

$$P(\text{Jenis Usaha} = \text{Property} | \text{Pendapatan Menurun}) = 2/32 = 0,06$$

$$P(\text{Jenis Usaha} = \text{Property} | \text{Pendapatan Stabil}) = 0/4 = 0$$

$$P(\text{Jenis Usaha} = \text{Online Shop} | \text{Pendapatan Menurun}) = 1/32 = 0,03$$

$$P(\text{Jenis Usaha} = \text{Online Shop} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

$$P(\text{Jenis Usaha} = \text{Cosmetic} | \text{Pendapatan Menurun}) = 1/32 = 0.03$$

$$P(\text{Jenis Usaha} = \text{Cosmetic} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

$$P(\text{Jenis Usaha} = \text{Fashion} | \text{Pendapatan Menurun}) = 0/32 = 0$$

$$P(\text{Jenis Usaha} = \text{Fashion} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

$$P(\text{Jenis Usaha} = \text{Wedding Accessories} | \text{Pendapatan Menurun}) = 1/32 = 0.03$$

$$P(\text{Jenis Usaha} = \text{Wedding Accessories} | \text{Pendapatan Stabil}) = 0/4 = 0$$

$$P(\text{Jenis Usaha} = \text{Parfum} | \text{Pendapatan Menurun}) = 0/32 = 0$$

$$P(\text{Jenis Usaha} = \text{Parfum} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

$$P(\text{Jenis Usaha} = \text{Cleaning Shoes} | \text{Pendapatan Menurun}) = 1/32 = 0.03$$

$$P(\text{Jenis Usaha} = \text{Cleaning Shoes} | \text{Pendapatan Stabil}) = 0/4 = 0$$

Probabilitas atribut Penghasilan:

$$P(\text{Penghasilan} = \text{Rendah} | \text{Pendapatan Menurun}) = 12/32 = 0.375$$

$$P(\text{Penghasilan} = \text{Rendah} | \text{Pendapatan Stabil}) = 2/4 = 0.5$$

$$P(\text{Penghasilan} = \text{Sedang} | \text{Pendapatan Menurun}) = 9/32 = 0.281$$

$$P(\text{Penghasilan} = \text{Sedang} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

$$P(\text{Penghasilan} = \text{Tinggi} | \text{Pendapatan Menurun}) = 11/32 = 0.343$$

$$P(\text{Penghasilan} = \text{Tinggi} | \text{Pendapatan Stabil}) = 1/4 = 0.25$$

3. Kalikan hasil variabel Pendapatan menurun dan Stabil menggunakan data baru [8] berikut:

Status = Menikah

Pendidikan = Perguruan Tinggi

Jenis Usaha = Fashion

Penghasilan = Sedang

$$\begin{aligned} &P(\text{Status} = \text{Menikah} | \text{Pendapatan menurun}) \times \\ &P(\text{Pendidikan} = \text{Perguruan Tinggi} | \text{Pendapatan} \\ &\text{menurun}) \times P(\text{Jenis Usaha} = \text{Fashion} | \text{Pendapatan} \\ &\text{Menurun}) \times P(\text{Penghasilan} = \text{Sedang} | \text{Pendapatan} \\ &\text{Menurun}) \times P(\text{Menurun}) \\ &= 0.375 \times 0.343 \times 0 \times 0.281 \times 0.89 \\ &= 0 \end{aligned}$$

$$\begin{aligned} &P(\text{Status} = \text{Menikah} | \text{Pendapatan Stabil}) \times P(\text{Pendidikan} \\ &= \text{Perguruan Tinggi} | \text{Pendapatan Stabil}) \times P(\text{Jenis} \\ &\text{Usaha} = \text{Fashion} | \text{Pendapatan Stabil}) \times P(\text{Penghasilan} = \\ &\text{Sedang} | \text{Pendapatan Stabil}) \times P(\text{Stabil}) \\ &= 0.25 \times 0.5 \times 0.25 \times 0.25 \times 0.11 \\ &= 0,00085 \end{aligned}$$

Karena nilai probabilitas Pendapatan Menurun lebih kecil dari nilai probabilitas Pendapatan Stabil, maka dapat disimpulkan bahwa data baru diatas termasuk dalam kategori Pendapatan Stabil.

B. Pengujian Metode Klasifikasi Menggunakan Data Testing

Hasil proses klasifikasi metode *naïve bayes* menggunakan teknik *split validation* pada data testing yang berjumlah 4 *record* dihasilkan proses klasifikasinya pada Tabel Klasifikasi *Data testing* :

Tabel 3
Klasifikasi *Data testing*

No	Umur	Status	Pendidikan	Jenis Usaha	Penghasilan	Actual Class	Predict Class
1	24	Lajang	Perguruan Tinggi	Cosmetik	Tinggi	Menurun	Stabil
2	19	Lajang	SMA/SMK	T-shirt	Sedang	Menurun	Menurun
3	22	Menikah	SMA/SMK	Makanan Olshop.	Tinggi	Menurun	Menurun
4	24	Lajang	Perguruan Tinggi	menjual luxury watch	Tinggi	Stabil	Stabil

*Dalam menentukan *Predict Class* pada tabel data testing diatas, menggunakan bantuan perhitungan *Naive Bayes* menggunakan Microsoft Excel.

Dari hasil proses klasifikasi yang disajikan pada tabel Klasifikasi *data testing* maka, dapat dikonversi kedalam bentuk tabel *confusion matrix*. Berikut Bentuk Tabel Pengujian *confusion matrix*:

TABEL IV
PENGUJIAN *CONFUSION MATRIX*

Classification	Predicted Class	
	Menurun	Stabil
Actual: Menurun	2 (TP)	1 (FN)
Actual: Stabil	0 (FP)	1 (TN)

TP (*True Positive*) → Merupakan data positif yang diprediksi benar.
FP (*False Positive*) → Merupakan data negatif namun diprediksi sebagai

data positif.

FN (*False Negative*) → Merupakan data positif namun diprediksi sebagai

data negatif.

TN (*True Negative*) → Merupakan data negatif yang diprediksi benar.
Berdasarkan tabel *confusion matrix* diatas maka kinerja dari

penggunaan metode klasifikasi *naive bayes* dapat diukur dengan menghitung nilai *accuracy*, *precision* dan *recall*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$= \frac{2+1}{2+1+0+1} \times 100\%$$

$$= \frac{3}{4} \times 100\%$$

$$= 0.75 \times 100\%$$

$$= 75\%$$

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$= \frac{2}{2+0} \times 100\%$$

$$= \frac{2}{2} \times 100\%$$

$$= 1 \times 100\%$$

$$= 100\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$= \frac{2}{2+1} \times 100\%$$

$$= \frac{2}{3} \times 100\%$$

$$= 0.667 \times 100\%$$

III. SIMPULAN

A. Simpulan

- 1) Sistem klasifikasi pedagang kaki lima dan pelaku usaha *online* yang dilakukan melalui pengisian *Kuesioner* dengan *Google form* yang terbukti rata-rata mengalami penurunan, hal tersebut dapat dibuktikan dari pengujian manual metode *Naive Bayes* menggunakan Atribut pendapatan pada tabel *data training*. Berdasarkan hasil diperoleh, probabilitas pendapatan menurun 89% dan probabilitas pendapatan stabil 11%.
- 2) Hasil pengujian *confusion matrix* dengan teknik split validasi, penggunaan metode klasifikasi *Naive Bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi 75%, berarti tingkat kedekatan nilai prediksi dengan nilai aktual pada tabel *data testing* termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 100%, berarti tingkat keakuratan antara data yang diminta dengan hasil prediksi pada tabel *data testing*. *Recall* sebesar 66.7%, berarti keberhasilan model dalam menemukan kembali sebuah informasi. Hal ini menyatakan bahwa sebagian besar klasifikasi pendapatan sudah sesuai. Kesalahan dalam klasifikasi juga tergantung pada dataset yang kita ujikan. jumlah dataset yang dipilih sebagai *data testing* mempengaruhi nilai *accuracy*, *precision* dan *recall*. Banyak ataupun sedikit atribut yang digunakan mempengaruhi jumlah data yang diklasifikasi pada setiap kategori. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi yang dibangun dapat dimaksimalkan sebagai bahan masukan bagi pengambil keputusan.

B. Saran

- 1) Mengingat nilai akurasi masih berada pada angka 75%, maka masih sangat dimungkinkan untuk dapat dilakukan penelitian selanjutnya, untuk meningkatkan nilai akurasi dengan menambahkan fitur seleksi atau penggunaan Model klasifikasi yang lain dan Pengujian dari Dataset metode *Naive Bayes* dengan *Rapid Miner* atau *Software* lainnya.
- 2) Pada penelitian lain diharapkan dapat digunakan Dataset dalam jumlah yang lebih besar atau dengan sejumlah variable lainnya guna meningkatkan performa dari metode yang digunakan.

UCAPAN TERIMA KASIH

Terwujudnya hasil artikel ini tidak lepas dari bantuan berbagai pihak yang telah bersedia menjadi narasumber untuk membantu mengisi *kuesioner* melalui *Google Form*. Terima kasih kepada para pihak yang sudah membantu mewujudkan artikel ini.

REFERENSI

- [1]. M. D. Vijayanti, and I. G. W. M Yasa, "Pengaruh lama usaha dan modal terhadap pendapatan Efisiensi usaha Pedagang sembako," *E-Jurnal Ekonomi Pembangunan*, Vol. 5, No 12, Desember 2016. [Online Serial]. Available: <https://media.neliti.com/media/publications/165217-ID-pengaruh-lama-usaha-dan-modal-terhadap-p.pdf> [Accessed April 03, 2020].
- [2]. H. Naparin, "Klasifikasi Peminatan Siswa SMA menggunakan Metode Naive Bayes," *Jurnal Systemic*, Vol. 2, No. 1, Agustus 2016. [Abstract]. Available: <https://media.neliti.com/media/publications/165217-ID-pengaruh-lama-usaha-dan-modal-terhadap-p.pdf> [Accessed April 04, 2020].
- [3]. F. Handayani, and F. S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *Jurnal Teknik Elektro*, Vol. 7, No. 1, January-June 2015. [Abstract]. Available: <https://journal.unnes.ac.id/nju/index.php/jte/article/view/8585/5674>, [Accessed April 07, 2020]
- [4]. F. Ariani, Amir, N. Alam, K. Rizal, "Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode Naive Bayes," *ILKOM Jurnal Komputer Dan Informatika*, Vol. 20, No. 2, September 2018. [Abstract]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/paradigma>, [Accessed April 07, 2020]
- [5]. Bustami, "Penerapan Algoritma Naive Bayes Untuk Menklasifikasi Data Nasabah Asuransi" *Jurnal Penelitian Teknik Informatika*" Vol. 8, No. 1, Januari 2014. [Abstract]. Available: <https://www.neliti.com/id/publications/102632/penerapan-algoritma-naive-bayes-untuk-mengklasifikasi-data-nasabah-asuransi>, [Accessed April 07, 2020]
- [6]. A. H. Nasrullah, "Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *ILKOM Jurnal Ilmiah*, Vol. 10, No. 2, August 2018. [Abstract]. Available: <https://media.neliti.com/media/publications/269212-penerapan-metode-c45-untuk-klasifikasi-m-f57b91ca.pdf>, [Accessed April 07, 2020]
- [7]. Galih, "Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, Vol. 2, No. 1, Januari 2019. [Abstract]. Available: <https://docplayer.info/storage/101/151725934/151725934.pdf>, [Accessed April 07, 2020]
- [8]. A. Lasarudin, and Purwanto, "Klasifikasi Pengaduan Masyarakat Menggunakan Naive Bayes Berbasis Seleksi Atribut Information Gain," *Jurnal Teknologi Informasi*, Vol 14, No 2, July 2018. [Abstract]. Available: <http://research.pps.dinus.ac.id/index.php/Cyberku/article/view/65>, [Accessed April 07, 2020]
- [9]. M. F. Fibrianda, and A. Bhawiyuga, "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naive Bayes Dan Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol 2, No. 9, September 2018. [Online Serial]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2559> [Accessed April 03, 2020].
- [10]. H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *ILKOM Jurnal Ilmiah*, Vol. 10, No. 2, August 2018. [Abstract]. Available: <https://media.neliti.com/media/publications/269202-klasifikasi-masyarakat-miskin-menggunaka-9d119a32.pdf>, [Accessed April 07, 2020]