



PENGGUNAAN NAIVE BAYES UNTUK PREDIKSI MINAT PASAR KEWIRAUSAHAAN BERDASARKAN DATA *MICROBLOGGING*

Dwi Meilvinasvita^{1)*}, Nabila Mufida²⁾, Muhammad Arhami³⁾, Salahuddin⁴⁾

¹⁾ Manajemen Keuangan Sektor Publik, Politeknik Negeri Lhokseumawe, Aceh
^{2,3,4)} Teknik Informatika, Politeknik Negeri Lhokseumawe, Aceh
email: dwimeilvinasvita@pnl.ac.id¹⁾, nabilamuff@gmail.com²⁾,
muhhammad.arhami@pnl.ac.id³⁾, salahuddintik@pnl.ac.id⁴⁾

Abstrak

Masalah awal dalam berwirausaha biasanya para pelaku wirausaha sulit untuk memahami permintaan serta kebutuhan calon konsumen. Ketidakpastian pasar dan perubahan permintaan membuat para wirausahawan tidak dapat memprediksi permintaan dan kebutuhan pasar. Tujuan penelitian ini melakukan evaluasi yang kuat pada kumpulan informasi yang didapat dari data *microblogging* untuk memperkirakan variabel pasar. Dengan membuat perancang sistem prediksi minat pasar wirausaha melalui pemanfaatan platform *microblogging twitter* menggunakan metode *naïve bayes*, sistem ini dapat menghasilkan model prediksi pasar secara otomatis melalui pemanfaatan data *microblogging* populer. Metode *naïve bayes* digunakan untuk mengklasifikasikan data untuk minat pasar tertentu hasil prediksi ini diperkirakan akan menjadi cara efektif untuk meningkatkan kinerja dan keunggulan bersaing dalam berwirausaha. Kumpulan dataset diambil dari salah satu platform *microblogging* yaitu aplikasi twitter dengan lima kata kunci dari cabang bisnis pakaian diantaranya celana, jaket, mukena, tunik dan kebaya. Dataset yang di ambil dibagi menjadi dua, data latih dan data uji. Data latih digunakan untuk membuat model prediksi sedangkan data uji dijadikan sebagai data evaluasi perancangan model. Setelah melakukan pengujian, akurasi sistem yang dihasilkan model sebesar 87%. Model yang dibuat menggunakan metode *naïve bayes* dikategorikan sangat baik dalam melakukan prediksi. Hasil akurasi prediksi dapat dilihat dalam bentuk berita populer sebagai acuan dalam menyediakan produk sesuai dengan permintaan dan kebutuhan calon konsumen.

Kata Kunci : Prediksi Pasar, *microblogging*, Naïve Bayes.

Abstract

The initial problem in entrepreneurship is that it is usually difficult for entrepreneurs to understand the demands and needs of potential consumers. Market uncertainty and changes in demand make entrepreneurs unable to predict market demand and needs. The purpose of this study is to conduct a robust evaluation of the collection of information obtained from microblogging data to estimate market variables. By designing a predictive system for entrepreneurial market interest through the use of the Twitter microblogging platform using the naive Bayes method, this system can generate market prediction models automatically through the use of popular microblogging data. The naive Bayes method is used to classify data for certain market interests. This prediction result is expected to be an effective way to improve performance and competitive advantage in entrepreneurship. The dataset is taken from one of the microblogging platforms, namely the twitter application with five keywords



from the clothing business branch including pants, jackets, mukena, tunic and kebaya. The dataset taken is divided into two, training data and test data. The training data is used to make prediction models, while the test data are used as evaluation data for the design of the model. After testing, the accuracy of the system generated by the model is 87%. The model made using the naive Bayes method is categorized as very good at making predictions. The results of prediction accuracy can be seen in the form of popular news as a reference in providing products according to the demands and needs of potential consumers.

Keywords: Market Prediction, microblogging, Naive Bayes.

PENDAHULUAN

Ilmu pengetahuan dan Teknologi terus berkembang seiring dengan peradaban manusia yang terus mengalami perubahan sesuai dengan tingkat kemampuan berpikir manusia yang semakin tinggi. Banyak inovasi-inovasi baru yang telah dihasilkan dan telah memberikan banyak manfaat yang positif bagi kehidupan manusia terutama dalam membantu aktifitas-aktifitas kesehariannya.

Terobosan teknologi selalu menjadi harapan baru dalam menyelesaikan persoalan-persoalan yang dihadapi oleh manusia dan penyelesaian tersebut dapat menjadi cara dalam meningkatkan efisiensi, kreatifitas, produktifitas, inovasi dan profesionalitas dalam dimensi bisnis dan ketenagakerjaan.

Badan Pusat Statistik (BPS) pada Februari 2021 mencata bahwa ada 205,36 juta jiwa penduduk Indonesia yang masuk kategori penduduk usia kerja. Dari angka tersebut ada 139,81 juta jiwa atau 68,08% merupakan kategori angkatan kerja. Dan sisanya 65,55 juta jiwa masuk kategori bukan angkatan kerja dimana salah satunya adalah ibu rumah tangga yang tidak bekerja. Rincian lebih lanjut dari angka angkatan kerja tersebut yaitu 131,06 juta jiwa memiliki status bekerja dan 8,75 juta jiwa memiliki status pengangguran terbuka [1].

Dampak dari pengangguran teknologi mungkin berkonsentrasi pada pekerja

berpenghasilan rendah. Sehingga akan ada urgensi untuk menghasilkan pelaku wirausaha dan peningkatan pelatihan pekerja agar dapat melakukan pekerjaan yang terampil dan bergaji tinggi [2]

Masalah awal dalam berwirausaha biasanya para pelaku wirausaha sulit untuk memahami permintaan serta kebutuhan calon konsumen. Ketidakpastian pasar dan perubahan permintaan membuat para wirausahawan tidak dapat memprediksi permintaan dan kebutuhan pasar. Bagaimana membuat calon konsumen bisa mendapatkan apa yang mereka butuhkan adalah tugas utama.

Saat ini Ledakan informasi sangat banyak dihasilkan dan diperbarui. Informasi ini, dapat dimanfaatkan untuk melakukan prediksi pasar secara otomatis. sebagai salah satu jalan pemenuhan kebutuhan calon konsumen. Volume data yang sangat besar bisa didapat dari berbagai sumber, salah satunya melalui *platform microblogging*. *Microblogging* adalah pesan atau postingan singkat yang diupload masyarakat di media sosial. Keberagaman dan kecepatan perkembangan data tersebut, dapat dijadikan data analisis yang memungkinkan ekstraksi data menjadi informasi berharga [3].

Penelitian ini mengusulkan evaluasi yang kuat pada kumpulan informasi yang didapat dari data *microblogging* untuk memperkirakan variabel pasar. Membuat



model prediksi pasar secara otomatis menggunakan konten data microblogging populer, diperkirakan akan menjadi cara efektif untuk meningkatkan kinerja dan keunggulan bersaing dalam berwirausaha. Kumpulan data yang dimaksud diambil dari media sosial jeni *twitter* sebagai salah satu *platform microblogging*. Data yang di ambil akan di klasifikasikan dengan metode *naïve bayes* untuk mendapatkan model dan hasil prediksi [4].

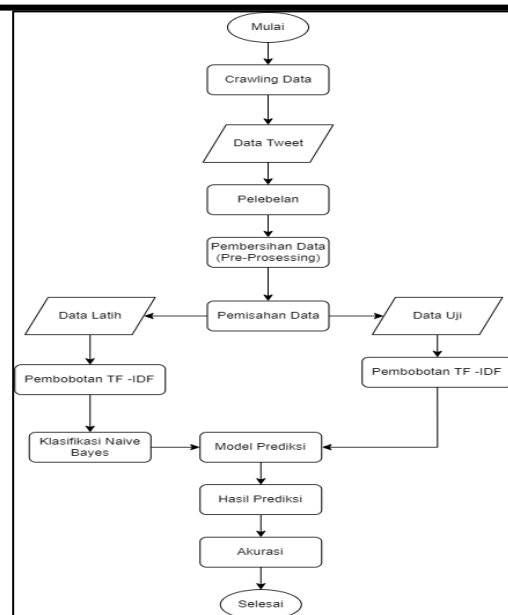
Hasil akurasi prediksi dapat dilihat dalam bentuk berita populer sebagai acuan dalam menciptakan produk sesuai dengan permintaan dan kebutuhan calon konsumen. Prediksi ini nantinya diharapkan dapat memudahkan para pelaku wirausaha untuk menciptakan pelayanan sesuai dengan keinginan calon konsumen serta meningkatkan kepuasan calon konsumen [5][6].

METODE PENELITIAN

A. Arsitektur Umum

Gambar 1 menjelaskan arsitektur umum perancangan sistem. proses dimulai dari scraping untuk memperoleh data *hoaks* dan tidak *hoaks* Masing-masing data kemudian dilakukan proses *preprocessing*. Tahap *preprocessing* adalah tahapan untuk membersihkan dan mempersiapkan data yang belum terstruktur menjadi data terstruktur melalui beberapa tahap, yaitu, *case folding*, *data cleaning*, *filtering*, *stemming*, dan *tokenizing*. *Dataset* selanjutnya dipisah dalam bentuk *data training* (data latih) dan *data test* (data uji).

Data latih akan digunakan melatih algoritma, dan data uji digunakan untuk menguji, dan mengukur kinerja model machine learning dan mengevaluasinya untuk selanjutnya merancang model machine learning dan implementasinya akan dilakukan pada website[7]



Gambar 1. Alur Perancangan sistem

B. Dataset

Dataset yang digunakan dalam penelitian berjumlah 1500 data. *Dataset* terbagi menjadi 5 kategori yaitu: celana, jaket, mukena, tunik dan kebaya. Masing masing-data berjumlah 300 data yang diambil dari salah satu *platform microblogging Twitter*.

C. Preprocessing

Proses *preprocessing* dilakukan untuk membersihkan data yang masih belum terstruktur sehingga data tersebut siap digunakan untuk merancang model machine learning. Tahapan-tahapan yang dilakukan meliputi:

Alur perancangan sistem pada gambar 1 dapat diterangkan seperti berikut:

1. Crawling data

Web crawling merupakan teknik pengumpulan data untuk indeksasi informasi halaman dengan menggunakan URL (*Uniform Resource Locator*) yang menyertakan API (*Application*



-
- Programming Interface*) untuk mining dataset yang ukurannya lebih besar. *Data mining* yang ada pada API sifatnya *open source* seperti yang disediakan pada *platform Twitter*.
2. *Pre processing*
Sebelum diolah menggunakan algoritma *Naïve Bayes* data *Twett* akan terlebih dahulu melewati proses *Pre-Processing*. Format *tweet* yang dikumpulkan perlu menjalani pra-pemrosesan untuk pelabelan data dan meningkatkan klasifikasi hasil. pra-pemrosesan berfungsi untuk mengekstraksi semua konten yang relevan dan fitur yang diperlukan dari *tweet* dan menjatuhkan konten yang tidak relevan[8].
 3. *Transform case*
Transform case memiliki tujuan utama mengubah semua huruf dalam dokumen menjadi huruf kecil. Beberapa cara menggunakan *Transform case* diantaranya: mengubah text menjadi huruf kecil, menghapus angka, tanda baca, *whitepace* (karakter kosong).
 4. *Filter Stopword*
Kata-kata penting dari hasil token diambil dengan menggunakan algoritma stoplist yaitu dilakukan dengan membuang kata kurang penting atau *wordlist* yaitu menyimpan kata penting. *Stopword* merupakan kata-kata umum yang biasanya muncul dalam jumlah besar dan tidak memiliki makna. Contoh *stopword* dalam bahasa Indonesia adalah “dan”, “yang”, “di”, “dari”, dan lain-lain.
 5. *Tokenize* (memisahkan kalimat menjadi per kata)
Tahapan ini adalah proses memisahkan kalimat menjadi perkata yang nantinya di representasikan dalam angka yang dimengerti mesin.
 6. *Stemming*
Tahapan ini adalah proses mengubah kata berimbuhan menjadi kata dasar.
 7. *Pemisahaan data*
Setelah melakukan pebelan proses selanjutnya adalah pemisahan data menjadi 2 kelompok data yaitu data latih dan data uji. Data latih (*training*) dan data uji (*test*) dibagi dengan rasio 70:30. Pembagian data ini dilakukan untuk mendapatkan hasil performasi yang baik.
 8. *Pembobotan*
Pada proses ini dilakukan pembobotan pada data *tweet* yang telah di dapat dengan tujuan memperoleh rating untuk kata-kata yang telah di dapatkan untuk dilakukan pengklasifikasian pada proses selanjutnya.
 9. *Klasifikasi naïve bayes*
Pada proses ini dilakukan klasifikasi data telah didapat untuk mendapatkan hasil prediksi dari *tweet* yang di ambil.
 10. *Pelebelan Otomatis*
Pada proses ini data uji di label secara otomatis oleh sistem setelah dilakukan klasifikasi.
 11. *Akurasi*
Akurasi yang dilakukan untuk melihat performansi dari prediksi ini adalah melalui pengukuran nilai akurasi, *precision*, dan *recall*.
 12. *Hasil*
Hasil yang diharapkan adalah sistem dapat memprediksikan trend pasar yang ada dikalangan masyarakat

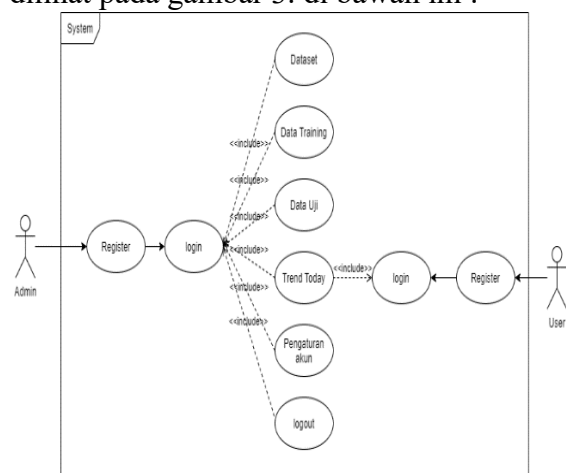


saat ini melalui data *microblogging twitter*.

Pengguna dapat menginput teks kalimat atau mengunggah dokumen, serta melihat hasil klasifikasi keduanya. *Admin* bertugas untuk merancang *dataset* dan merancang model *machine learning*.

D. Use Case Diagram

Perancangan *Use Case diagram* dapat dilihat pada gambar 3. di bawah ini :

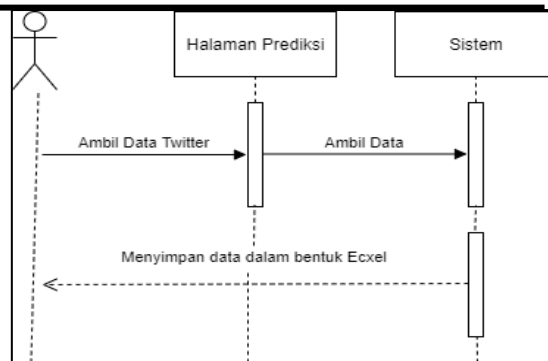


Gambar 2. Use Case Diagram

Gambar 2 menunjukkan bahwa ada 2 aktor dalam system ini, yaitu *user* (pengguna) dan *admin*. Langkah pertama adalah pengguna dapat melakukan *register* dan *login* untuk melihat trend di dalam sistem. dan *Admin* dapat mengakses beberapa menu seperti menu *dataset*, menu data uji, *menu trend*, menu pengaturan akun dan menu *logout*.

E. Sequence Diagram

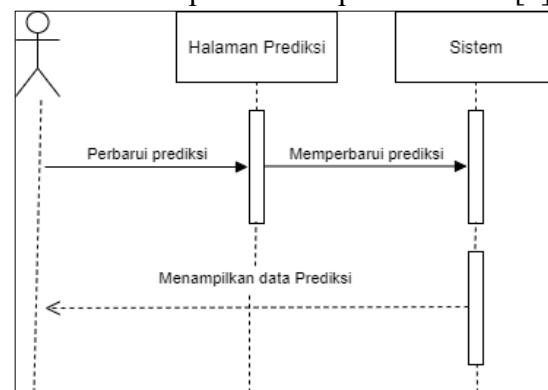
Gambar 3 merupakan diagram sequens dari proses pengambilan data *twitter*. Saat admin menekan tombol ambil data twitter maka akan dikirimkan pesan ambil data ke sistem. Seanjutnya sistem langsung menyimpan data *tweet* dalam bentuk *exel* ke perakat lokal *user*.



Gambar 3. Diagram Sequens Perbarui Prediksi

F. Diagram Sequens Perbarui Trend

Gambar 4 merupakan diagram *sequens* dari proses memperbarui prediksi *twitter*. Saat *admin* menekan tombol ambil data *twitter* maka akan dikirimkan pesan perbarui prediksi ke sistem. Selanjutnya sistem menampilkan data prediksi baru [9].



Gambar 4. Diagram Sequens Ambil Data Twitter

HASIL DAN PEMBAHASAN

A. Implementasi Proses

Implementasi proses menjabarkan prosedur dan tahapan yang terlibat berdasarkan perancangan model machine learning yang telah dijabarkan.

1. Web Scraping

Proses *web scraping* dijalankan dengan menggunakan API Twitter



berdasarkan kata yang ingin di ambil total terdapat 1500 data.

	id_str	text	DeteksiTopik
0	1.551843e+18	convomf rekomendasi celana kulot yang nyaman b...	celana
1	1.551842e+18	convomf rekomendas celana hits masa kini #cela...	celana
2	1.551840e+18	pengen beli jaket winter yang tebal banget! #Mi...	jaket
3	1.551830e+18	rekomendasi jaket oversized murah dan terbaru...	jaket
4	1.551859e+18	RT @ohmybeautybank: lagi cari muke trevel supa...	mukena
5	1.551859e+18	RT @ohmybeautybank: [tb] Mukena travel ALIF MO...	mukena
6	1.551864e+18	Jait kebaya woi .cari cari kebaya buat wisuda...	kebaya
7	1.551864e+18	metawinlovesme please kekompas beli kebaya ...	kebaya

Gambar 5. Diagram Sequens Ambil Data Twitter

2. Data Cleaning

Data dibersihkan dari segala karakter yang tidak berguna dan dapat mempengaruhi performa seperti angka, emoji, tanda baca, link, simbol. Hasil proses data cleaning ditunjukkan pada Gambar 6.

remove_html	tweet_clean	tweet
convomf rekomendasi celana kulot yang nyaman b...	convomf, rekomendasi, celana, kulot, nyaman, b...	convomf rekomendasi celana kulot yang nyaman b...
convomf rekomendas celana hits masa kini #cela...	convomf, rekomendas, celana, hits, masa, kini, #cela...	convomf rekomendas celana hits masa kini #cela...
pengen beli jaket winter yang tebal banget! #Mi...	pengen, beli, jaket, winter, yang, tebal, banget!, #Mi...	pengen beli jaket winter yang tebal banget! #Mi...
rekomendasi jaket oversized murah dan terbaru...	rekomendasi, jaket, oversized, murah, dan, terbaru...	rekomendasi jaket oversized murah dan terbaru...
lagi cari muke trevel supa...	lagi, cari, muke, trevel, supa...	lagi cari muke trevel supa...
[tb] Mukena travel ALIF MODERN HEAR Juara b...	[tb], mukena, travel, alif, modern, hear, juara, b...	[tb] Mukena travel ALIF MODERN HEAR Juara b...
Jait kebaya woi .cari cari kebaya buat wisuda...	Jait, kebaya, woi, .cari, cari, kebaya, buat, wisuda...	Jait kebaya woi .cari cari kebaya buat wisuda...
metawinlovesme please kekompas beli kebaya ...	metawinlovesme, please, kekompas, beli, kebaya, ...	metawinlovesme please kekompas beli kebaya ...

Gambar 6. Hasil Data Cleaning

3. Labeling

Labeling dilakukan pada kalimat berdasarkan jenis 0= celana, 1=jeket, mukena, 2= mukena, 3 = tunik, 4 = kebaya.

```
df.DeteksiTopik.replace("celana",0, inplace = True)
df.DeteksiTopik.replace("jaket",1, inplace = True)
df.DeteksiTopik.replace("mukena",2, inplace = True)
df.DeteksiTopik.replace("tunik",3, inplace = True)
df.DeteksiTopik.replace("kebaya",4, inplace = True)
```

Gambar 7. Code Proses Labeling Data

```
def prediksi(tweet):
    text = clean_tweets(tweet)
    text = remove_punct(text)
    tfidf_vektor = tfidf_vectorizer.transform([text])
    pred = joblib_model.predict( tfidf_vektor)
    if pred == 0 :
        DeteksiTopik= 'Topik dari kalimat adalah celana'
    elif pred == 1 :
        DeteksiTopik= 'Topik dari kalimat adalah jaket'
    elif pred == 2 :
        DeteksiTopik= 'Topik dari kalimat adalah mukena'
    elif pred == 3 :
        DeteksiTopik= 'Topik dari kalimat adalah tunik'
    else :
        DeteksiTopik= 'Topik dari kalimat adalah kebaya'
    return DeteksiTopik
```

Gambar 8. Code Labeling Otomatis Data Uji

4. TF-Idf dan Klasifikasi *Naïve Bayes*
Klasifikasi yang digunakan dalam artikel ini adalah klasifikasi Multinomial *Naïve Bayes* yang diimplementasikan pada data-data *Tweet* yang terboboti yang telah dilakukan pada tahap ekstraksi fitur sebelumnya. Tahap klasifikasi untuk pengerjaan aplikasi menggunakan *library* yang ada dalam bahasa pemrograman Python yang dikenal dengan sebutan *scikit-learn*. Sebelum proses klasifikasi dilakukan, langkah pertama yang dilakukan adalah melakukan akses file data cuitan yang sudah melewati *preprocessing* sebagai data training yang menggunakan *library* *pandas*. Kode program yang digunakan seperti pada gambar 9. Selanjutnya dilakukan ekstraksi fitur dan klasifikasi sekaligus. Data uji yang digunakan untuk tahap klasifikasi ini sebesar 20% dari jumlah data latih yang dipilih secara acak oleh sistem. Pengujian dilakukan dengan klasifikasi sebanyak 5 (lima) kali dengan sebarang data uji yang bervariasi. Variasi Data uji yang bervariasi diperoleh dengan memberi nilai berbeda pada variabel *random_state* pada pengujian ini,



Nilai *random_state* yang digunakan adalah 0.1-0.9.

```

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

[ ] tfidf_vectorizer = TfidfVectorizer().fit(text)
tfidf_text = tfidf_vectorizer.transform(text)

from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(tfidf_text, label, test_size = 0.3, random_state=123)

from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report
model = MultinomialNB().fit(x_train, y_train)
pred_nb = model.predict(x_test)
print(classification_report(pred_nb, y_test))
from sklearn.metrics import confusion_matrix
confusion_matrix(pred_nb, y_test)

precision recall f1-score support
0 1.00 1.00 1.00 13
1 1.00 0.92 0.96 13
2 1.00 1.00 1.00 16
3 0.00 1.00 0.00 7
4 1.00 1.00 1.00 13

accuracy 0.90
macro avg 0.97 0.98 0.98 62
weighted avg 0.99 0.98 0.98 62

array([[1, 0, 0, 0, 0],
       [0, 12, 0, 1, 0],
       [0, 0, 16, 0, 0],
       [0, 0, 0, 7, 0],
       [0, 0, 0, 0, 13]])
    
```

Gambar 9. Code Proses Tf-idf dan Naive Bayes

Gambar 9 merupakan beberapa kalimat untuk mendeteksi topik dari kalimat.

```

[36] prediksi('ayah beli celana dimana')
    'Topik dari kalimat adalah celana'

[37] prediksi('wisuda pengen pake kebaya cantik')
    'Topik dari kalimat adalah kebaya'
    
```

Gambar 10. Hasil Contoh Prediksi

5. Pada kalimat 1 “ayah beli celana dimana” berhasil di deteksi topik dari kalimat adalah celana. Dan pada kalimat 2 ”wisuda pengen kebaya cantik ” berhasil di deteksi topik dari kalimat adalah kebaya. Pengujian dan Evaluasi

TABEL 1. Hasil Pengujian Sistem

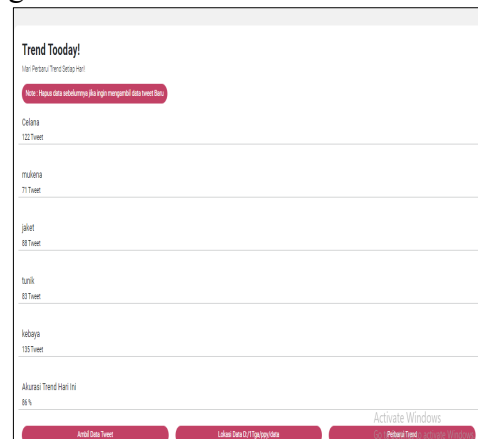
Data Latih	Data Uji	Precision	Recall	F1-Score	akurasi
10%	90%	64%	60%	59%	70%
20%	80%	68%	83%	64%	73%
30%	70%	68%	84%	69%	72%
40%	60%	79%	89%	82%	83%
50%	50%	81%	90%	83%	84%
60%	40%	83%	91%	85%	85%
70%	30%	81%	88%	83%	83%
80%	20%	83%	89%	85%	87%
90%	10%	86%	89%	86%	86%

6. pengujian dilakukan untuk menguji bagaimana performa sistem dalam menghasilkan prediksi. dalam hal ini pengujian dibuat untuk 10 skenario seperti pada tabel 1. Berdasarkan tabel 1 dapat dilihat bagaimana kenaikan nilai akurasi pada setiap scenario. Semakin banyak data yang digunakan sebagai data training maka semakin tinggi pula nilai akurasi yang di hasilkan. Begitu pula dengan nilai *recall* dan *precision*. Nilai ini digunakan untuk melihat seberapa besar kecocokan dan keberhasilan sistem dalam mendeteksi topik sesuai dengan label yang telah di berikan.

B. Implementasi Sistem

Implementasi sistem menjabarkan implementasi dari perancangan sistem.

1. Implementasi halaman *trend admin*
Implementasi halaman *trend admin* seperti yang ditunjukkan pada gambar 11.



Gambar 11. Halaman Trend Admin

Trend admin merupakan halaman untuk menampilkan trend dan akurasi dari pakaian setiap hari nya. Data trend di perbarui oleh *admin* setiap harinya dengan mengambil



data dari *twitter*, membuat model prediksi dan memperbarui *trend*

2. Implementasi Halaman *Trend User*
Trend User merupakan halaman untuk menampilkan *trend* dan akurasi dari pakaian setiap hari nya yang dapat dilihat oleh *user* dan mentor.

Trend Today!	
Lihat Trend Hari Ini Untuk Meningkatkan Perkembangan Usaha Mu!	
Celana	122 Tweet
Mukena	71 Tweet
Jaket	88 Tweet
Tunik	83 Tweet
Kebaya	135 Tweet
Akurasi Trend Hari Ini	
86 %	

Gambar 12. Halaman *Trend User*

Pada halaman *trend user* terdapat informasi topik yang paling sering dibicarakan dengan melihat jumlah *tweet* tertinggi. Pada gambar 12. dapat dilihat *tweet* yang paling sering dibicarakan adalah topik kebaya. Selanjutnya topik celana, jaket, mukena, dan topik yang paling jarang dibicarakan adalah topik kebaya.

SIMPULAN

Setelah melakukan perancangan dan pengujian pada sistem prediksi minat pasar melalui pemanfaatan *microblogging twitter* menggunakan *naïve bayes*:

1. Metode *naïve bayes* dapat digunakan untuk membantu mengklasifikasikan data *tweet* dalam memprediksikan minat pasar wirausaha.
2. Sistem mampu mendeteksi topik dengan melakukan pembobotan

menggunakan metode TF-IDF dan melakukan klasifikasi menggunakan metode *naïve bayes*. Dalam proses deteksi sistem memperoleh nilai *recall* dan presision sebesar 83% dan 89% sehingga sistem dinilai mampu mendeteksi kalimat dengan baik.

3. prediksi minat pasar melalui pemanfaatan *microblogging twitter* menggunakan *naïve bayes* menghasilkan akurasi terbesar pada skenario 80:20 dengan nilai akurasi sebesar 87%.

DAFTAR PUSTAKA

- [1] Edy Arsyad, "Pengangguran di Indonesia Capai 8,75 Juta Orang," <https://fajar.co.id/>, May 28, 2021.
- [2] G. E. Marchant, Y. A. Stevens, and J. M. Hennessy, "Technology, Unemployment & Policy Options: Navigating the Transition to a Better World," *J Evol Technol*, vol. 24, 2014.
- [3] Nadiyah Rahmalia, "Microblogging: Definisi, Keunggulan, dan Beragam Platformnya," <https://glints.com/>, Feb. 11, 2021.
- [4] A. Hadikusuma, A. Fat, H. Hidayat, and D. M. Ramadhani, "Paper Review: Data Mining Twitter," 2014.
- [5] A. Desiani, R. Primartha, M. Arhami, and O. Orsalan, "Naive Bayes classifier for infant weight prediction of hypertension mother," in *Journal of Physics: Conference Series*, Aug. 2019, vol. 1282, no. 1. doi: 10.1088/1742-6596/1282/1/012005.
- [6] W. Agastya et al., "Pemetaan Emosi Dominan pada Kalimat Majemuk



-
- Bahasa Indonesia Menggunakan Multinomial Naïve Bayes (Mapping Dominant Emotion in Indonesian Compound Sentences Using Multinomial Naïve Bayes),” 2020.
- [7] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, p. 427, Oct. 2018, doi: 10.25126/jtiik.201854773.
- [8] A. S. Nayak and A. P. Kanive, “Survey on Pre-Processing Techniques for Text Mining,” *International Journal Of Engineering And Computer Science*, Jun. 2016, doi: 10.18535/ijecs/v5i6.25.
- [9] P. Sulistyorini, “Pemodelan Visual dengan Menggunakan UML dan Rational Rose,” *Jurnal Teknologi Informasi DINAMIK Volume XIV*, No.1, Januari 2009 : 23-29, vol. XIV, Jan. 2009.