VISUALISASI DATA HASIL KLASIFIKASI NAÏVE BAYES DENGAN MATPLOTLIB PADA PYTHON

Siti Mujilahwati[*]
Jurusan Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan
Jl. Veteran 53 A Lamongan – Jawa Timur.
[*]Email: moedjee@gmail.com

Abstrak

Data apabila disajikan dalam bentuk visual maka akan mudah dibaca. Terutama data yang memiliki pengaruh pada pengambilan keputusan. Misalkan naik turunnya omset penjualan, stok produk dan lain sebagainya. Hampir semua aplikasi data menyediakan fitur visualisasi data berupa grafik, misalkan excel. python juga menyediakan library untuk menvisualisasikan sebuah data dalam bentuk grafik yaitu matplotlib. Pada penelitian ini membahas beberapa fungsi plot yang digunakan untuk menvisualisasikan data hasil klasifikasi topik abstrak skripsi dengan menggunakan metode Naïve Bayes. Dari hasil klasifikasi tersebut divisualisasikan dalam grafik heatmap, Lini plot, Scatter plot dan Histogram plot. Dari ke empat grafik yang diperoleh dapat memberikan informasi pergerakan antara data faktual dengan data prediksi ke kelompok masing-masing kelas.

***Kata kunci**: visualisasi, python, matplotlib, Klasifikasi*

1. PENDAHULUAN

Seiring pertumbuhan data yang semakin hari semakin bertambah, maka akan mengakibatkan penumpukan data yang mubadzir jika pemilik data tidak dapat melakukan pengolahan data dengan benar. Peran data sangat besar untuk kemajuan sebuah bisnis, data dapat diolah supaya dapat mengetahui trend yang akan terjadi di masa depan. Saat ini banyak cara untuk melakukan analisis data pada penerapan data science ataupun data analysis. Visualisasi data adalah menggambarkan secara nyata data yang bergerak baik dalam bentuk tabel, bar chart, pie chart, line graph, map dan diagram.

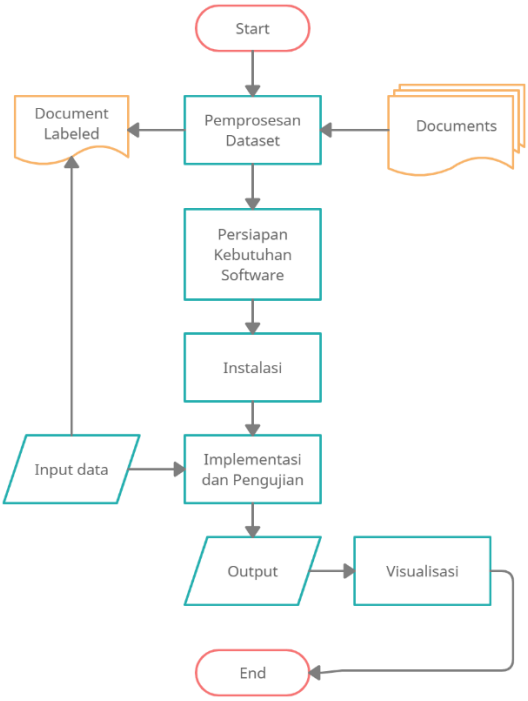
Untuk melakukan visualisasi data juga beragam cara, seperti yang telah dilakukan oleh peneliti sebelumnya memanfaatkan Exploratory Data Analysis untuk melakukan visualisasi data bunuh diri diseluruh negara di dunia. Dengan melakukan visualisasi tersebut Irwan dapat memberikan gambaran pada negara mana saja dan tahun berapa data lonjakan naik turun tingkat bunuh diri (Setiawan et al., 2021). Selain Irwan peneliti sebelumnya dari salah satu universitas swasta di Jakarta juga melakukan data visualisasi untuk melihat kinerja dan performance perguruan tinggi yang dipimpinnya, tujuan dari visualisasi data ini adalah untuk melakukan pengambilan keuptusan pada kebijakan kebijakan pimpinan(Waruwu & Wulandari, 2020). Dessy dan Johan juga telah mencoba melakukan visuali sasi data penjualan pada suatu PT dengan metode visual data mining (VDM) dan Exploratory Data Analysis (EDA), dari hasil penelitiannya maka PT dapat melihat produk mana saja yang efisien dan tidak(Aryanti & Setiawan, 2019). Dedy Hartama juga melakukan Teknik visualisasi dengan model Tableau Big Data untuk menganalisis data akademik pada data status mahasiswa dengan tujuan mengetahui secara cepat keadaan perkembangan database akademik(Hartama, 2018).

Klasifikasi merupakan sebuah Teknik pengelompokkan pada data yang memiliki kemiripan atau kesamaan dalam satu kelompok, penelitian ini melakukan pengelompokkan tema skripsi berdasarkan dokumen abstrak. Klasifikasi yang dilakukan dengan memanfaatkan Naïve Bayes Multinomial. Dari hasil klasifikasi yang diperoleh akan disajikan dalam bentuk visualisasi dengan memanfaatkan library matplotlib pada python. Klasifikasi adalah salah satu bagian dari data mining, klafikasi memiliki arti pengelompokkan. Teknik klasifikasi merupakan proses pembelajaran terarah (Supervise Learning)(Darujati & Gumelar, n.d.). Hasil pengelompokkan akan didasarkan dari hasil data pelatihan atau data terdahulu. Salah satu algoritma yang dapat memberikan hasil terbaik untuk Teknik klasifikasi adalah Naïve Bayes(Asril & Kamila, 2019; Kalokasari et al., 2017; Wijaya & Santoso, n.d.)

Dari beberapa uraian di atas maka dapat disimpulkan bahawa melakukan visualisasi data salah satu cara agar pergerakan data dapat secara cepat dibaca, dianalisis dan selanjutnya dapat cepat dilakukan pengambilan keputusan. Dari artikel atau penelitian yang dikutip di atas yang membedakan dari penelitian ini adalah cara melakaukan visualisasi dan juga data yang akan dilakukan visualisasi merupakan data hasil klasifikasi bukan data yang belum diproses atau data tersimpan. Data klassifikasi secara real akan bergerak sehingga melakukan visualisasi akan sangat membantu pembacaan hasil klasifikasi topik skripsi.

2. METODE

Pada penelitian data yang digunakan adalah data skripsi mahasiswa Teknik informatika pada lulusan tahun 2019 sebanyak 126 dokumen. Dari dokumen tersebut hanya data abstrak yang diambil tiap masing- masing mahasiswa. Kerangka penelitian ini dapat digambarkan pada Gambar 1 berikut ini.



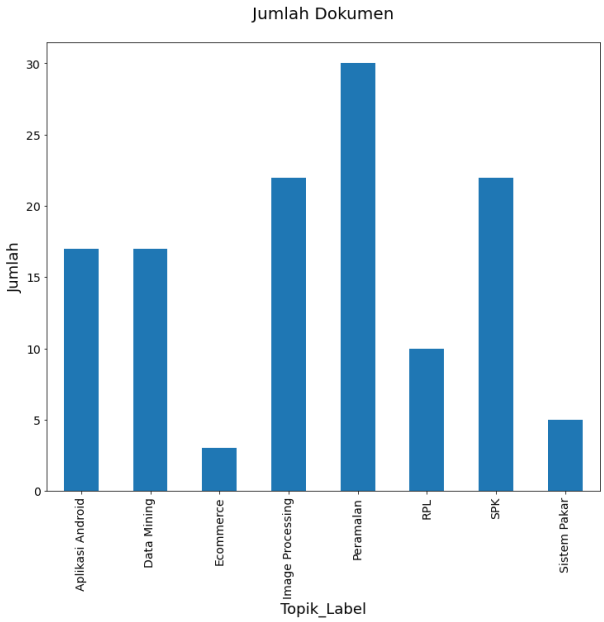
Gambar 1. Kerangka Alur Penelitian

Pertama yang dilakukan adalah pengambilan data, data dilakukan pengecekan kelayakan data, selanjutnya data dilakukan analisis secara manual untuk melakukan kelas pada topik yang sudah ditentukan yaitu ada 6 kelas dan pada proses persiapan data diperoleh data sebagai berikut.

Tabel 1. Dataset Dokumen Abstrak

No	Kelas	Jumlah Dokumen
1	Aplikasi Android	17
2	Data Mining	17
3	RPL	10
4	SPK	5
5	Pengolahan Citra Digital	22
6	Sistem Pakar	5
7	e-Commerce	3
8	Peramalan	30
	Total	126

ISBN 978-602-52386-5-9



Gambar 2. Sebaran Data pada Class

Setelah dataset telah disiapkan selanjutnya adalah melakukan persiapan software yang akan digunakan. Platform yang digunakan untuk implementasi pada penelitian ini adalah Jupyter Notebook. Ada beberapa library yang perlu diinstall pada penelitian dini pada python yaitu :

1. Matplotlib
2. Sklearn
3. Numpy
4. Pandas
5. NBClussifier

Setelah library sudah siap maka selanjutnya adalah melakukan persipan data di python dengan cara sebagai berikut.

```
def load_data():
   data =  pd.read_excel('dataset20.xlsx')
 return data
```

Selanjutnya melakukan split data training dan testing.

```
In [15]: abstrak['Kelas'].value_counts()

Out[15]: Aplikasi_Android            38
         Data_Mining                 19
         RPL                         16
         SPK                         14
         Pengolahan_Citra_Digital     8
         Sistem_Pakar                 8
         Name: Kelas, dtype: int64

In [16]: train_size = int(len(abstrak) * .8)
         print ("Train size: %d" % train_size)
         print ("Test size: %d" % (len(abstrak) - train_size))

         Train size: 82
         Test size: 21

In [17]: def train_test_split(abstrak, train_size):
             train = abstrak[:train_size]
             test = abstrak[train_size:]
             return train, test
```

Gambar 3. Fungsi Split data Train dan Testing

Setelah data sudah terbagi 80% sebagai data latih dan 20% data test selanjutnya dilakukan token atau praproses dengan library keras. Selanjutnya dilakukan proses klasifikasi berdasarkan kelas

yang sudah ditentukan. Dan hasilnya akan dilakukan visualisasi dengan library matplotlib pada python.

3. HASIL DAN PEMBAHASAN

Hasil klasifikasi yang diperoleh pada penelitian ini pada data pengujian atau testing adalah test loss sebesar 1,0946 sedangkan test Accurasi sebesar 0,6190 atau 61,90%. Dengan penambahan optimasi tuning dengan optuna maka diperoleh nilai test accurasi 76,19. Maka akurasi optimal ini yang akan digambarkan atau divisualisasikan dan dibahas pada penelitian ini, dengan menggunakan matplotlib.

1. Import Library Matplotlib di python

```
In [ ]: import matplotlib.pyplot as plt
```

Jika belum terinstal, kita harus install terlebih dahulu dengan cara berikut.

```
C:\Users\Your Name>pip install matplotlib
```

Instalasi tentu melalui platform python.

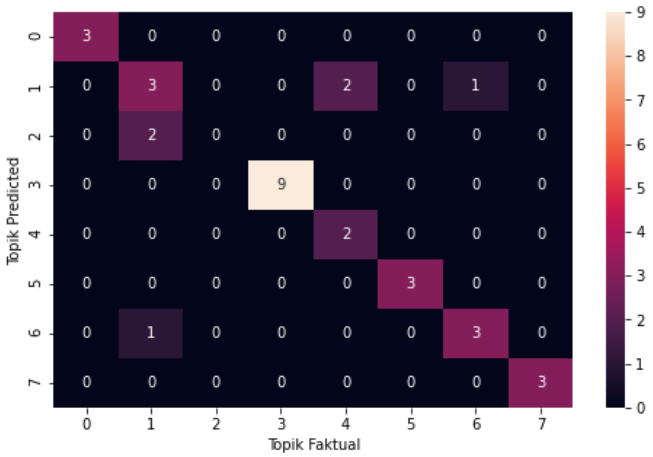
Berikutnya pada pembahasan artikel ini akan membahas 4 gambar visualisasi pada hasil klasifikasi yang diperoleh.

1) Heatmap

Grafik ini akan menggambarkna dengan jelas hasil klasifikasi antara data factual dengan data prediksi. Atau dapat benar dengan data salah. Fungsi yang dipakai pada penelitian ini adalah sebagai berikut

```
import seaborn as sns
import matplotlib.pyplot as plt
f, ax = plt.subplots(figsize=(8,5))
sns.heatmap(metrics.confusion_matrix(y_test, y_pred_class),
annot=True, fmt=".0f", ax=ax)
plt.xlabel("Topik Faktual")
plt.ylabel("Topik Predicted")
plt.show()
```

Gambar grafik yang diperoleh adalah sebagai Gambar 4 berikut.



Gambar 4. Grafik Heatmap Plot

ISBN 978-602-52386-5-9

2) Line

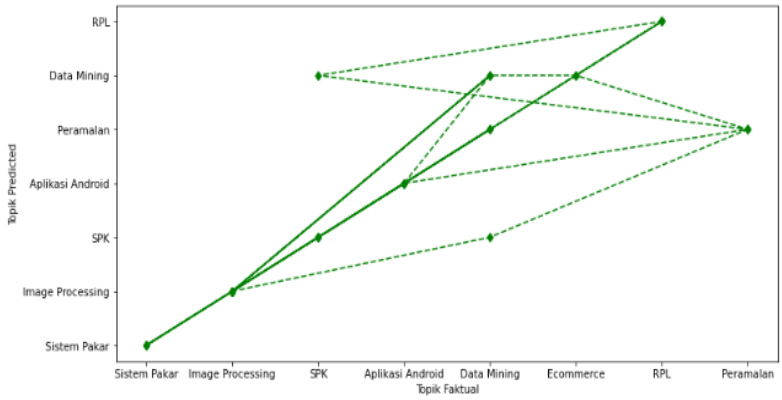
Plot ini merupakan rangkaian dari titik titik data yang saling terhubung sehingga menggambarkan pergerakan data dan sering digunakan sebagai plot visualisasi dasar. Fungsi yang digunakan pada penelitian ini adalah seperti berikut.

```
import matplotlib.pyplot as plt
#line plot

plt.figure(figsize=(12,6))
plt.plot(y_test, y_pred_class,'g--d')

plt.xlabel("Topik Faktual")
plt.ylabel("Topik Predicted")
plt.show()
```

Hasil diagram Line plot



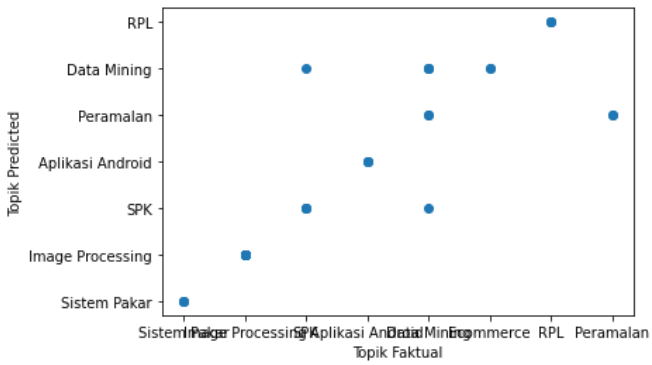
Gambar 5. Grafik Line Plot

3) Scatter

Grafik ini biasanya digunakan untuk memplot titik setiap pengamatan. Fungsi yang digunakan pada penelitian ini adalah sebagai berikut.

```
#Scatter Plot
plt.scatter(y_test, y_pred_class)
plt.xlabel("Topik Faktual")
plt.ylabel("Topik Predicted")
plt.show()
```

Hasil Gambar yang diperoleh dari fungsi ini adalah seperti Gambar 6 berikut.



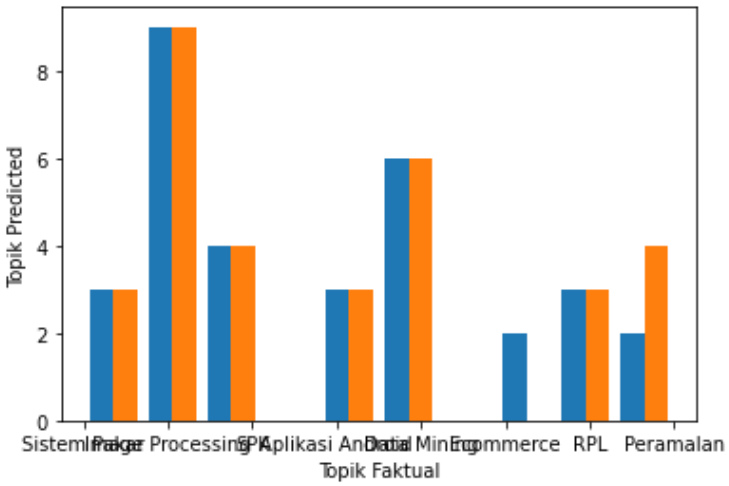
Gambar 6. Grafik Scatter Plot

4) Hstograms

Diagram ini salah satu diagram yang sering digunakan oleh peneliti untuk menggambarkan data. Fungsi yang pakai pada penelitian ini adalah sebagai berikut.

```
#histogram
x = np.random.normal(size=100)
plt.hist((y_test, y_pred_class), bins=10)
plt.xlabel("Topik Faktual")
plt.ylabel("Topik Predicted")
plt.show()
```

Gambar yang dihasilkan sebagai Gambar 7 berikut.



Gambar 7. Grafik Histogram Plot

Pada dasarnya untuk membuat sebuah gambar frafik atau plot grafik pada python sangatlah mudah. Hal tersebut dikarenakan, python memiliki fasilitas menampilkan grafik dari kumpulan data yang ada yang biasa disebut library. Kemampuan matplotlib untuk menggambarkan grafik bentuk 2D dan 3D harus memenuhi kreteria data berupa matrik 2 dimensi dan juga matrik 3 dimensi.

Hasil penelitian ini dengan menggambarkan hasil output atau klasifikasi berupa data 2 dimensi yaitu berupa data factual dan data prediksi. Dan dengan menggunakan fungsi yang ada di matplotlib mendaapatkan hasil yang cukup baik. Hanya saja belum dapat memberikan banyak variasi warna pada masing-masing kelas target klasifikasi.

4. KESIMPULAN

Dapat disimpulkan untuk hasil visialisasi hasil klasifikasi data abstrak sekripsi pada penelitian ini adalah sebagai berikut.

1) Python menyediakan library untuk membuat atau melakukan plot data berupa grafik
2) Plot yang digambarkan adalah hasil dari klasifikasi menggunakan algoritma Naïve Bayes Multinomial yang tersedia pada python
3) Gambar grafik masih belum dimaksimalkan untuk legenda dari sebaran data
4) Dengan hasil visualisasi maka pergerakan data missing atau tidak tepat pada kelasnya akan langsung dapat dilihat secara nyata berdasarkan jumlah data yang salah. Selain itu letak salah masuk kelas juga akan terlihat. Seperti pada Gambar 4 di atas terlihat secara jelas.

DAFTAR PUSTAKA

Aryanti, D., & Setiawan, J. (2019). Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018. *Ultima InfoSys*, *9*(2), 86–91. https://doi.org/10.31937/si.v9i2.991

Hartama, D. (2018). Analisa Visualisasi Data Akademik Menggunakan Tableau Big Data. *Jurasik*

ISBN 978-602-52386-5-9

(Jurnal Riset Sistem Informasi Dan Teknik Informatika), *3*(3), 46. https://doi.org/10.30645/jurasik.v3i0.65

Setiawan, I., Korespondensi, P., & Analysis, E. D. (2021). *Visualisasi dan analisis data bunuh diri*. *8*(3), 445–456. https://doi.org/10.25126/jtiik.202183391

Waruwu, L. M., & Wulandari, T. (2020). Perancangan Visualisasi Informasi Data Warehouse dan Dashboard System Data Perguruan Tinggi di Universitas Mercubuana Jakarta. *Jurnal Ilmu Teknik Dan Komputer*, *4*(2), 116–123.

Darujati, C., & Gumelar, A. B. (n.d.). PEMANFAATAN TEKNIK SUPERVISED UNTUK KLASIFIKASI TEKS BAHASA INDONESIA. 9

Asril, H., & Kamila, I. (2019). *Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor*. 10

Kalokasari, D. H., Shofi, I. M., & Setyaningrum, A. H. (2017). IMPLEMENTASI ALGORITMA MULTINOMIAL NAIVE BAYES CLASSIFIER PADA SISTEM KLASIFIKASI SURAT KELUAR (Studi Kasus: DISKOMINFO Kabupaten Tangerang). *JURNAL TEKNIK INFORMATIKA*, *10*(2), 109–118. https://doi.org/10.15408/jti.v10i2.6199

ijaya, A. P., & Santoso, H. A. (n.d.). *Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government*. *1*, 8