ACTIVITY PERTEMUAN 4

NAMA : MUHAMMAD TARMIDZI BARIQ

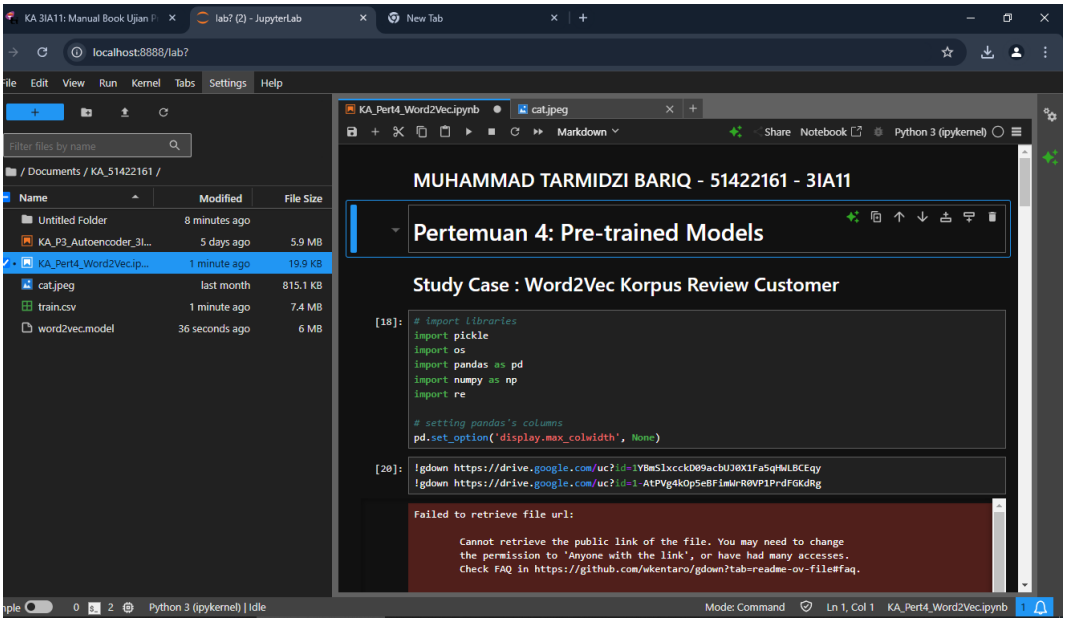
NPM : 51422161

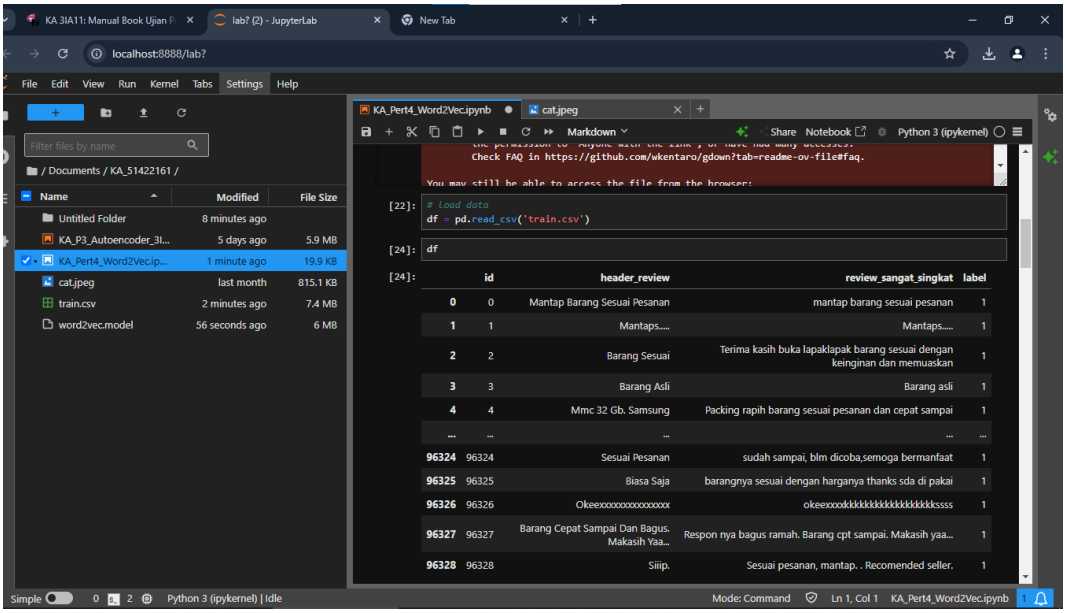
KELAS : 3IA11

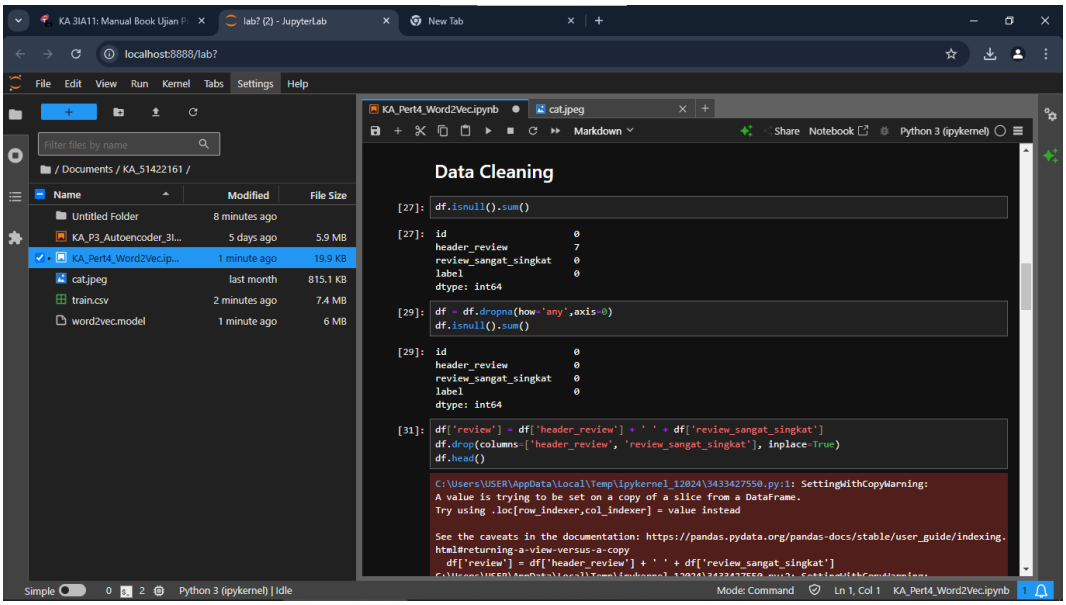
MATERI : Pre – trained Model

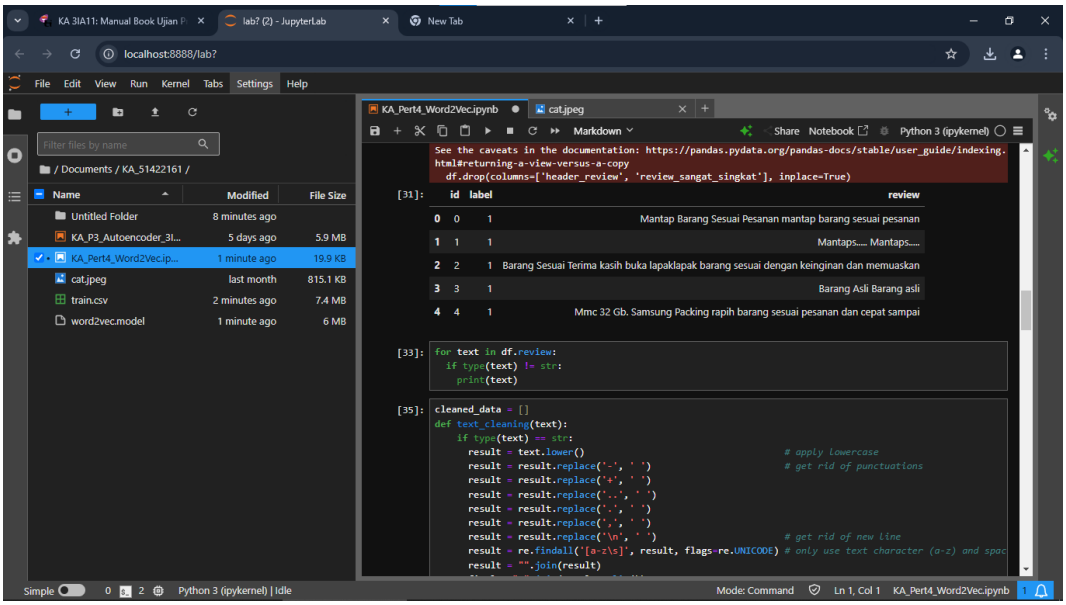
MATA PRAKTIKUM : KECERDASAN ARTIFICIAL

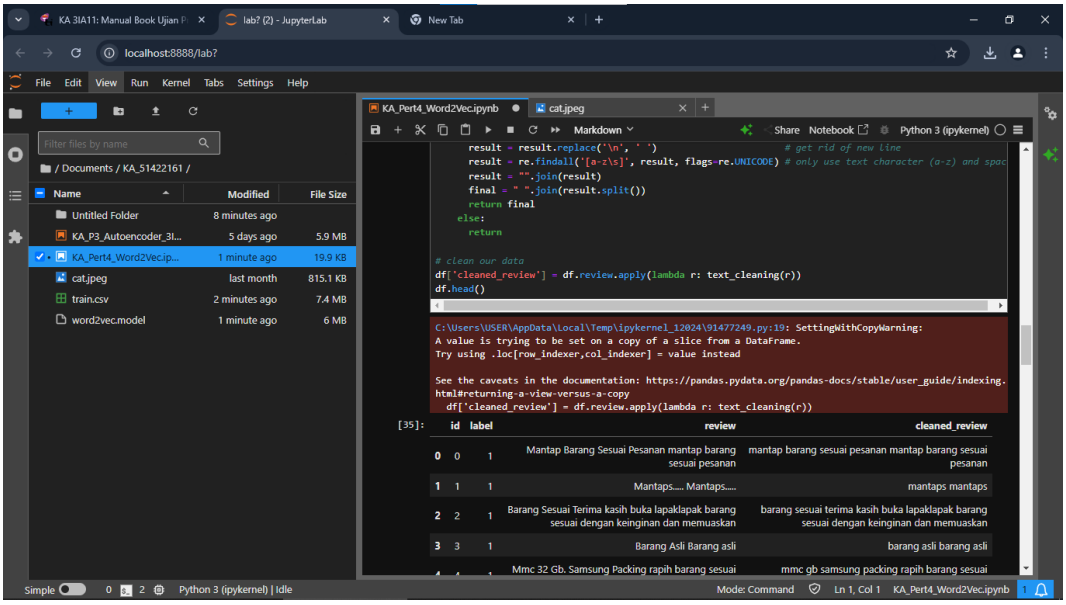
(Screenshoot langkah-langkah sesuai video pembelajaran dan jelaskan dengan ringkas)

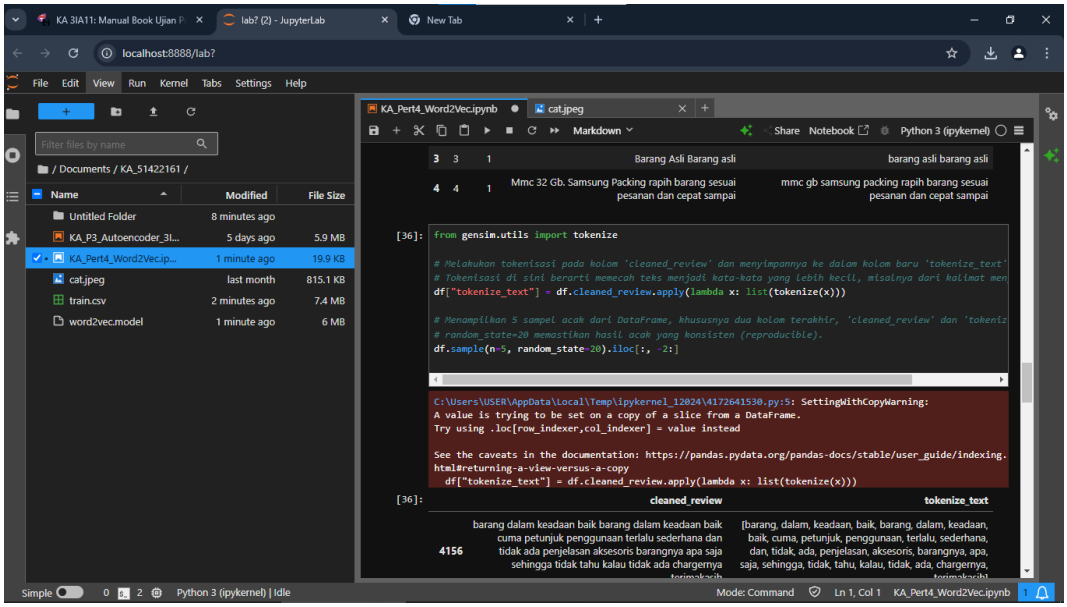


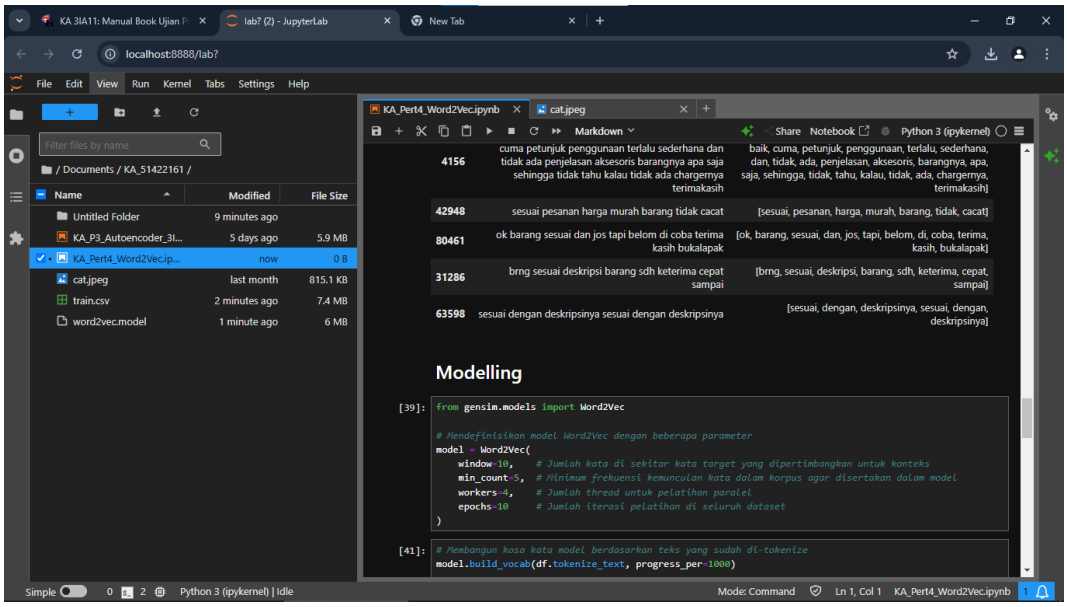

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook
Python 3 (ipykernel)
Check FAQ in https://github.com/wkentaro/gdown?tab=readme-ov-file#faq.
You may still be able to access the file from the browser:
[22]:
# Load data
df = pd.read_csv('train.csv')
[24]:
df
[24]:
id header_review review_sangat_singkat label
0 0 Mantap Barang Sesuai Pesanan mantap barang sesuai pesanan 1
1 1 Mantaps..... Mantaps..... 1
2 2 Barang Sesuai Terima kasih buka lapaklapak barang sesuai dengan keinginan dan memuaskan 1
3 3 Barang Asli Barang asli 1
4 4 Mmc 32 Gb. Samsung Packing rapih barang sesuai pesanan dan cepat sampai 1
... ...
96324 96324 Sesuai Pesanan sudah sampai, blm dicoba,semoga bermanfaat 1
96325 96325 Biasa Saja barangnya sesuai dengan harganya thanks sda di pakai 1
96326 96326 Okeexxxxxxxxxxxxxxx okeexxxxkkkkkkkkkkkkkkkkkkkkssss 1
96327 96327 Barang Cepat Sampai Dan Bagus, Makasih Yaa... Respon nya bagus ramah. Barang cpt sampai. Makasih yaa... 1
96328 96328 Siiip. Sesuai pesanan, mantap. . Recomended seller. 1
Filter files by name
/ Documents / KA_51422161 /
Name Modified File Size
Untitled Folder 8 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 1 minute ago 19.9 KB
cat.jpeg last month 815.1 KB
train.csv 2 minutes ago 7.4 MB
word2vec.model 56 seconds ago 6 MB
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb

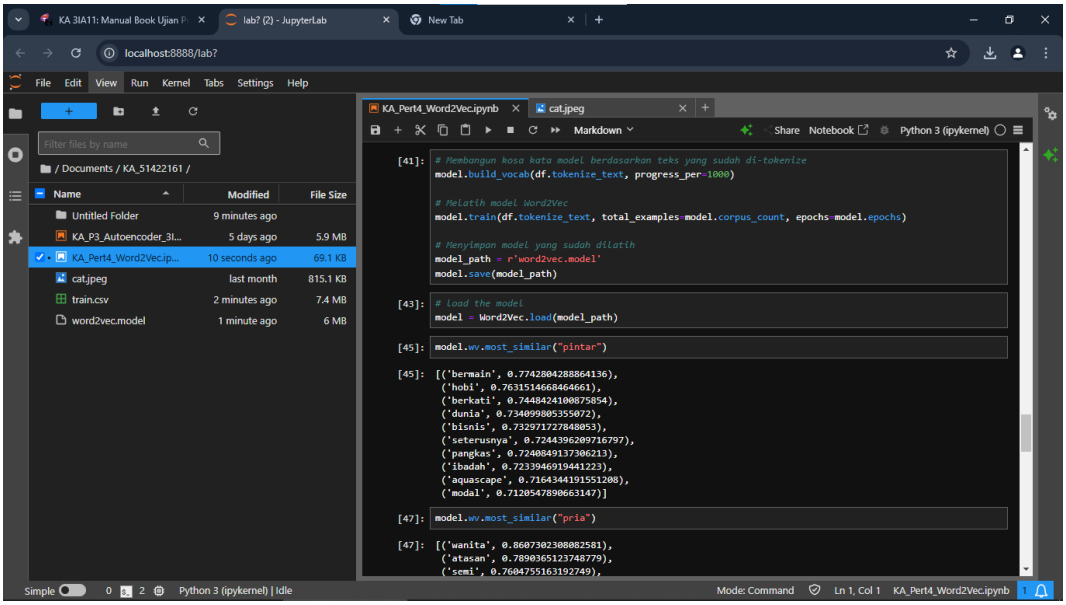

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook
Python 3 (ipykernel)
Data Cleaning
[27]:
df.isnull().sum()
[27]:
id 0
header_review 7
review_sangat_singkat 0
label 0
dtype: int64
[29]:
df = df.dropna(how='any',axis=0)
df.isnull().sum()
[29]:
id 0
header_review 0
review_sangat_singkat 0
label 0
dtype: int64
[31]:
df['review'] = df['header_review'] + ' ' + df['review_sangat_singkat']
df.drop(columns=['header_review', 'review_sangat_singkat'], inplace=True)
df.head()
C:\Users\USER\AppData\Local\Temp\ipykernel_12024\3433427550.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df['review'] = df['header_review'] + ' ' + df['review_sangat_singkat']
Filter files by name
/ Documents / KA_51422161 /
Name Modified File Size
Untitled Folder 8 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 1 minute ago 19.9 KB
cat.jpeg last month 815.1 KB
train.csv 2 minutes ago 7.4 MB
word2vec.model 1 minute ago 6 MB
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb

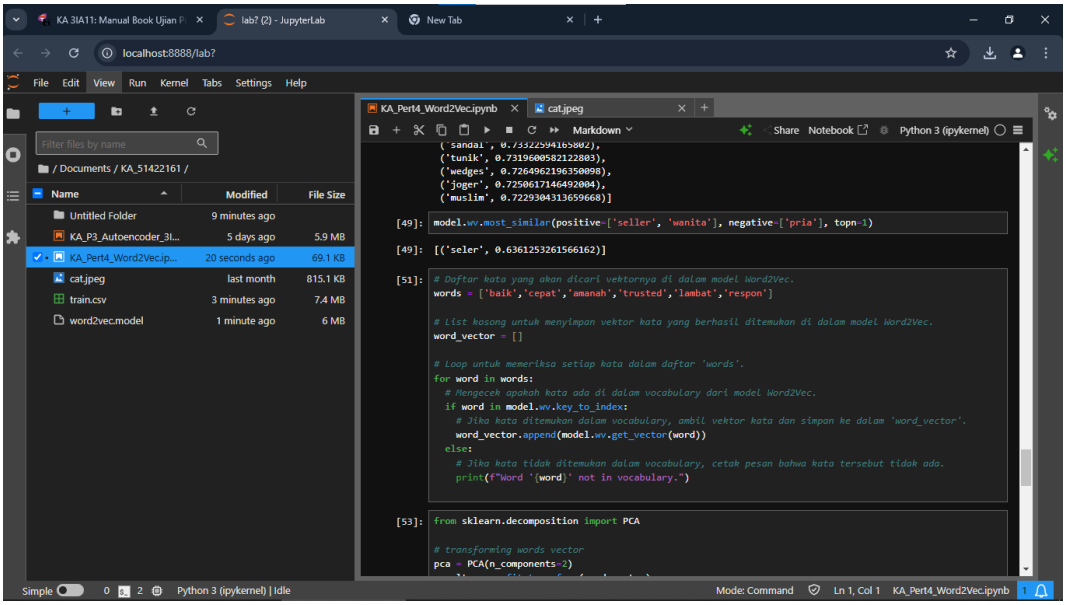

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook
Python 3 (ipykernel)
Filter files by name
/ Documents / KA_51422161 /
Name
Modified
File Size
Untitled Folder 8 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 1 minute ago 19.9 KB
cat.jpeg last month 815.1 KB
train.csv 2 minutes ago 7.4 MB
word2vec.model 1 minute ago 6 MB
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df.drop(columns=['header_review', 'review_sangat_singkat'], inplace=True)
[31]:
id label review
0 0 1 Mantap Barang Sesuai Pesanan mantap barang sesuai pesanan
1 1 1 Mantaps..... Mantaps.....
2 2 1 Barang Sesuai Terima kasih buka lapaklapak barang sesuai dengan keinginan dan memuaskan
3 3 1 Barang Asli Barang asli
4 4 1 Mmc 32 Gb. Samsung Packing rapih barang sesuai pesanan dan cepat sampai
[33]:
for text in df.review:
if type(text) != str:
print(text)
[35]:
cleaned_data = []
def text_cleaning(text):
if type(text) == str:
result = text.lower() # apply lowercase
result = result.replace('-', ' ') # get rid of punctuations
result = result.replace('+', ' ')
result = result.replace('..', ' ')
result = result.replace('.', ' ')
result = result.replace(',', ' ')
result = result.replace('\n', ' ') # get rid of new line
result = re.findall('[a-z\s]', result, flags=re.UNICODE) # only use text character (a-z) and spac
result = "".join(result)
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb

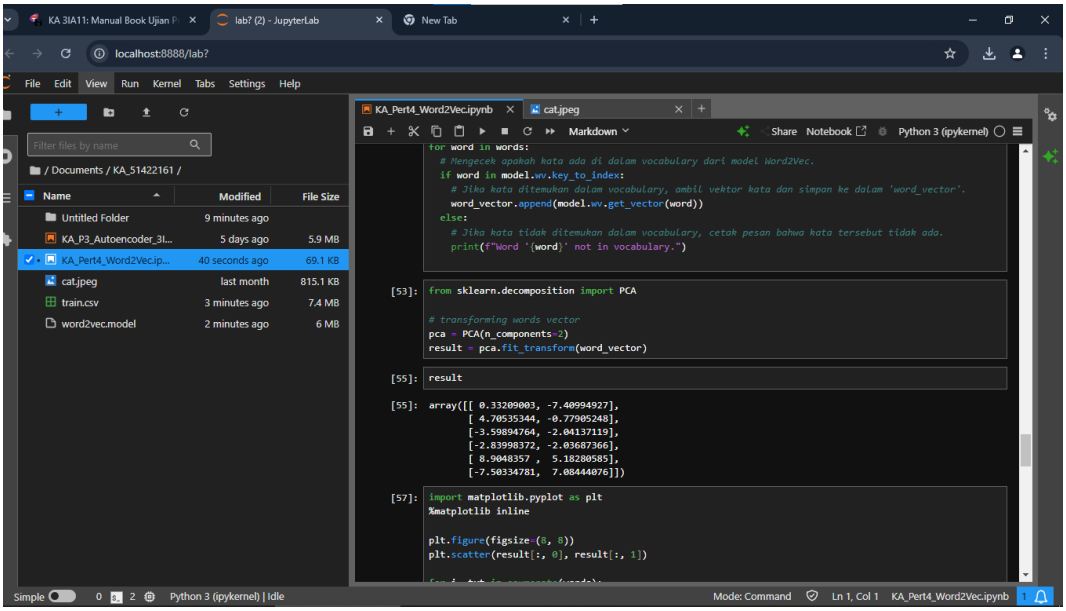

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook
Python 3 (ipykernel)
result = result.replace('\n', ' ') # get rid of new line
result = re.findall('[a-z\s]', result, flags=re.UNICODE) # only use text character (a-z) and spac
result = "".join(result)
final = " ".join(result.split())
return final
else:
return
clean our data
df['cleaned_review'] = df.review.apply(lambda r: text_cleaning(r))
df.head()
C:\Users\USER\AppData\Local\Temp\ipykernel_12024\91477249.py:19: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df['cleaned_review'] = df.review.apply(lambda r: text_cleaning(r))
[35]:
id label review cleaned_review
0 0 1 Mantap Barang Sesuai Pesanan mantap barang sesuai pesanan mantap barang sesuai pesanan mantap barang sesuai pesanan
1 1 1 Mantaps..... Mantaps..... mantaps mantaps
2 2 1 Barang Sesuai Terima kasih buka lapaklapak barang sesuai dengan keinginan dan memuaskan barang sesuai terima kasih buka lapaklapak barang sesuai dengan keinginan dan memuaskan
3 3 1 Barang Asli Barang asli barang asli barang asli
4 4 1 Mmc 32 Gb. Samsung Packing rapih barang sesuai mmc gb samsung packing rapih barang sesuai
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb

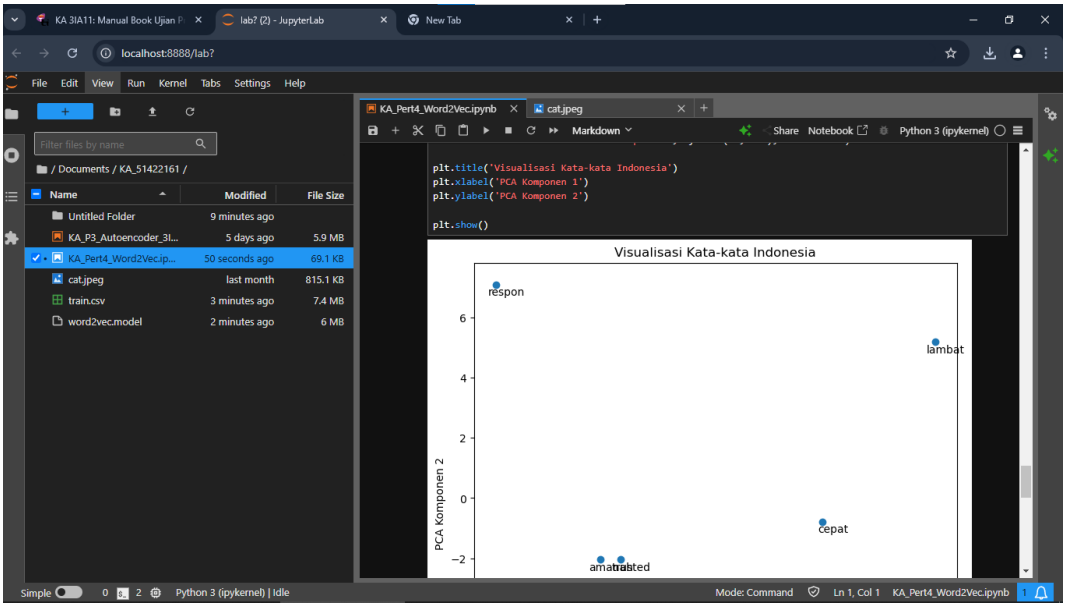

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook Python 3 (ipykernel)
/ Documents / KA_51422161 /
Name Modified File Size
Untitled Folder 8 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 1 minute ago 19.9 KB
cat.jpeg last month 815.1 KB
train.csv 2 minutes ago 7.4 MB
word2vec.model 1 minute ago 6 MB
3 3 1 Barang Asli Barang asli barang asli barang asli
4 4 1 Mmc 32 Gb. Samsung Packing rapih barang sesuai pesanan dan cepat sampai mmc gb samsung packing rapih barang sesuai pesanan dan cepat sampai
[36]:
from gensim.utils import tokenize
# Melakukan tokenisasi pada kolom 'cleaned_review' dan menyimpannya ke dalam kolom baru 'tokenize_text'
# Tokenisasi di sini berarti memecah teks menjadi kata-kata yang lebih kecil, misalnya dari kalimat men
df["tokenize_text"] = df.cleaned_review.apply(lambda x: list(tokenize(x)))
Menampilkan 5 sampel acak dari DataFrame, khususnya dua kolom terakhir, 'cleaned_review' dan 'tokeniz
random_state=20 memastikan hasil acak yang konsisten (reproducible).
df.sample(n=5, random_state=20).iloc[:, -2:]
C:\Users\USER\AppData\Local\Temp\ipykernel_12024\4172641530.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df["tokenize_text"] = df.cleaned_review.apply(lambda x: list(tokenize(x)))
[36]:
cleaned_review tokenize_text
4156 barang dalam keadaan baik barang dalam keadaan baik cuma petunjuk penggunaan terlalu sederhana dan tidak ada penjelasan aksesoris barangnya apa saja sehingga tidak tahu kalau tidak ada chargernya [barang, dalam, keadaan, baik, barang, dalam, keadaan, baik, cuma, petunjuk, penggunaan, terlalu, sederhana, dan, tidak, ada, penjelasan, aksesoris, barangnya, apa, saja, sehingga, tidak, tahu, kalau, tidak, ada, chargernya,
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb

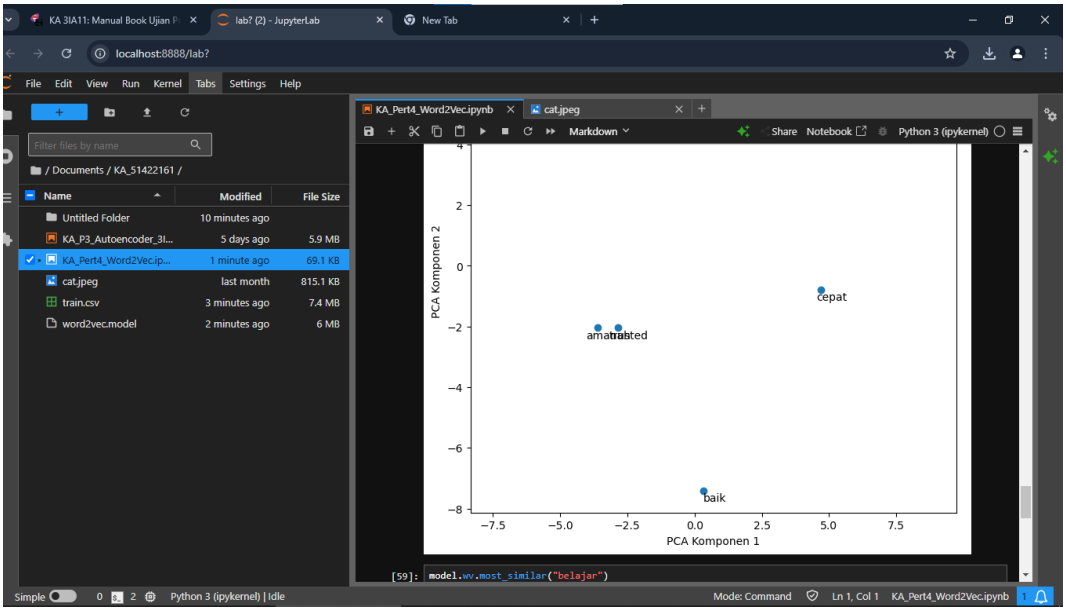

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook Python 3 (ipykernel)
/ Documents / KA_51422161 /
Name Modified File Size
Untitled Folder 9 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... now 0 B
cat.jpeg last month 815.1 KB
train.csv 2 minutes ago 7.4 MB
word2vec.model 1 minute ago 6 MB
4156 cuma petunjuk penggunaan terlalu sederhana dan tidak ada penjelasan aksesoris barangnya apa saja sehingga tidak tahu kalau tidak ada chargernya terimakasih baik, cuma, petunjuk, penggunaan, terlalu, sederhana, dan, tidak, ada, penjelasan, aksesoris, barangnya, apa, saja, sehingga, tidak, tahu, kalau, tidak, ada, chargernya, terimakasih]
42948 sesuai pesanan harga murah barang tidak cacat [sesuai, pesanan, harga, murah, barang, tidak, cacat]
80461 ok barang sesuai dan jos tapi belom di coba terima kasih bukalapak [ok, barang, sesuai, dan, jos, tapi, belom, di, coba, terima, kasih, bukalapak]
31286 brng sesuai deskripsi barang sdh keterima cepat sampai [brng, sesuai, deskripsi, barang, sdh, keterima, cepat, sampai]
63598 sesuai dengan deskripsinya sesuai dengan deskripsinya [sesuai, dengan, deskripsinya, sesuai, dengan, deskripsinya]
Modelling
[39]:
from gensim.models import Word2Vec
# Mendefinisikan model Word2Vec dengan beberapa parameter
model = Word2Vec(
window=10, # Jumlah kata di sekitar kata target yang dipertimbangkan untuk konteks
min_count=5, # Minimum frekuensi kemunculan kata dalam korpus agar disertakan dalam model
workers=4, # Jumlah thread untuk pelatihan paralel
epochs=10 # Jumlah iterasi pelatihan di seluruh dataset
)
[41]:
# Membangun kosa kata model berdasarkan teks yang sudah di-tokenize
model.build_vocab(df.tokenize_text, progress_per=1000)
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb

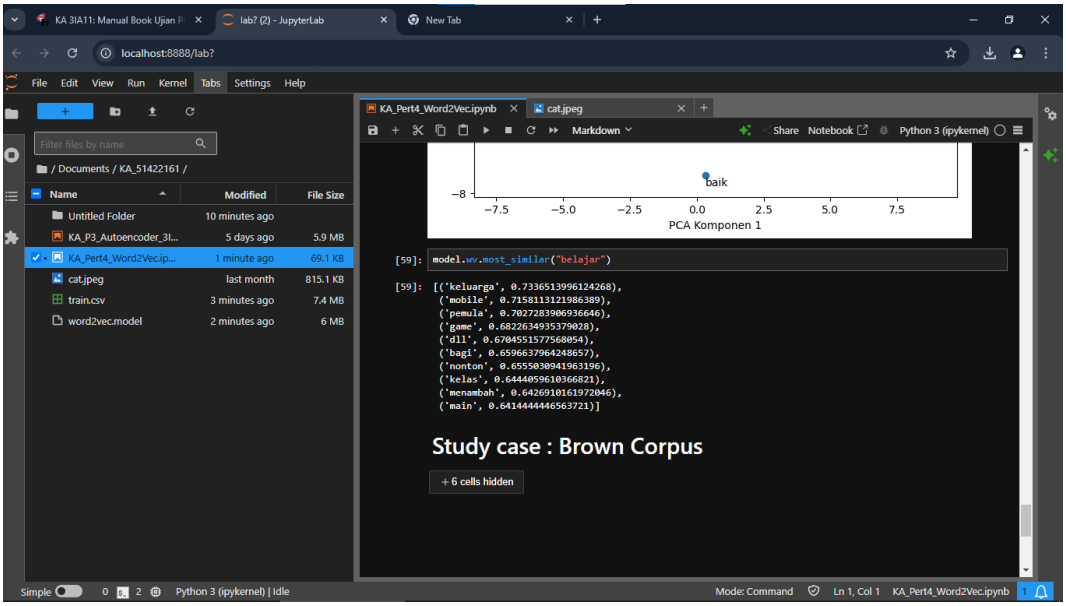

KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook
Python 3 (ipykernel)
Filter files by name
/ Documents / KA_51422161 /
Name
Modified
File Size
Untitled Folder 9 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 10 seconds ago 69.1 KB
cat.jpeg last month 815.1 KB
train.csv 2 minutes ago 7.4 MB
word2vec.model 1 minute ago 6 MB
[41]:
# Membangun kosa kata model berdasarkan teks yang sudah di-tokenize
model.build_vocab(df.tokenize_text, progress_per=1000)
# Melatih model Word2Vec
model.train(df.tokenize_text, total_examples=model.corpus_count, epochs=model.epochs)
# Menyimpan model yang sudah dilatih
model_path = r'word2vec.model'
model.save(model_path)
[43]:
# Load the model
model = Word2Vec.load(model_path)
[45]:
model.wv.most_similar("pintar")
[45]:
[('bermain', 0.7742804288864136),
('hobi', 0.7631514668464661),
('berkati', 0.7448424100875854),
('dunia', 0.7340998053550720),
('bisnis', 0.7329717278480530),
('seterusnya', 0.7244396209716797),
('pangkas', 0.7240849137306213),
('ibadah', 0.7233946919441223),
('aquascape', 0.7164344191551208),
('modal', 0.7128547890663147)]
[47]:
model.wv.most_similar("pria")
[47]:
[('wanita', 0.8607302308082581),
('atasan', 0.7890365123748779),
('semi', 0.7604755163192749),
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb 1


KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook
Python 3 (ipykernel)
Filter files by name
/ Documents / KA_51422161 /
Name
Modified
File Size
Untitled Folder 9 minutes ago
KA_P3_Autoencoder_3I... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 20 seconds ago 69.1 KB
cat.jpeg last month 815.1 KB
train.csv 3 minutes ago 7.4 MB
word2vec.model 1 minute ago 6 MB
('sandai', 0.7332594165802),
('tunik', 0.7319600582122803),
('wedges', 0.7264962196350098),
('joger', 0.7250617146492004),
('muslim', 0.7229304313659668)]
[49]:
model.wv.most_similar(positive=['seller', 'wanita'], negative=['pria'], topn=1)
[49]:
[('seler', 0.6361253261566162)]
[51]:
# Daftar kata yang akan dicari vektornya di dalam model Word2Vec.
words = ['baik','cepat','amanah','trusted','lambat','respon']
# List kosong untuk menyimpan vektor kata yang berhasil ditemukan di dalam model Word2Vec.
word_vector = []
# Loop untuk memeriksa setiap kata dalam daftar 'words'.
for word in words:
# Mengecek apakah kata ada di dalam vocabulary dari model Word2Vec.
if word in model.wv.key_to_index:
# Jika kata ditemukan dalam vocabulary, ambil vektor kata dan simpan ke dalam 'word_vector'.
word_vector.append(model.wv.get_vector(word))
else:
# Jika kata tidak ditemukan dalam vocabulary, cetak pesan bahwa kata tersebut tidak ada.
print(f"Word '{word}' not in vocabulary.")
[53]:
from sklearn.decomposition import PCA
# transforming words vector
pca = PCA(n_components=2)
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb 1


KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook Python 3 (ipykernel)
/ Documents / KA_51422161 /
Name Modified File Size
Untitled Folder 9 minutes ago
KA_P3_Autoencoder_3l... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 40 seconds ago 69.1 KB
cat.jpeg last month 815.1 KB
train.csv 3 minutes ago 7.4 MB
word2vec.model 2 minutes ago 6 MB
for word in words:
# Mengecek apakah kata ada di dalam vocabulary dari model Word2Vec.
if word in model.wv.key_to_index:
# Jika kata ditemukan dalam vocabulary, ambil vektor kata dan simpan ke dalam 'word_vector'.
word_vector.append(model.wv.get_vector(word))
else:
# Jika kata tidak ditemukan dalam vocabulary, cetak pesan bahwa kata tersebut tidak ada.
print(f"Word '{word}' not in vocabulary.")
[53]:
from sklearn.decomposition import PCA
# transforming words vector
pca = PCA(n_components=2)
result = pca.fit_transform(word_vector)
[55]:
result
[55]:
array([[0.3320903, -7.40994927],
[4.70535344, -0.77905248],
[-3.59894764, -2.04137119],
[-2.83998372, -2.03687366],
[8.9048357 , 5.18280585],
[-7.50334781, 7.08444076]])
[57]:
import matplotlib.pyplot as plt
%matplotlib inline
plt.figure(figsize=(8, 8))
plt.scatter(result[:, 0], result[:, 1])
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb


KA 3IA11: Manual Book Ujian P
lab? (2) - JupyterLab
New Tab
localhost:8888/lab?
File Edit View Run Kernel Tabs Settings Help
KA_Pert4_Word2Vec.ipynb
cat.jpeg
Markdown
Share Notebook Python 3 (ipykernel)
/ Documents / KA_51422161 /
Name Modified File Size
Untitled Folder 9 minutes ago
KA_P3_Autoencoder_3l... 5 days ago 5.9 MB
KA_Pert4_Word2Vec.ip... 50 seconds ago 69.1 KB
cat.jpeg last month 815.1 KB
train.csv 3 minutes ago 7.4 MB
word2vec.model 2 minutes ago 6 MB
plt.title('Visualisasi Kata-kata Indonesia')
plt.xlabel('PCA Komponen 1')
plt.ylabel('PCA Komponen 2')
plt.show()
Visualisasi Kata-kata Indonesia
respon
lambat
cepat
amat
rabted
PCA Komponen 2
6
4
2
0
−2
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 KA_Pert4_Word2Vec.ipynb





2. Jelaskan tujuan dari praktikum tadi

Efisiensi Waktu dan Sumber Daya, Karena model sudah dilatih sebelumnya, kita tidak perlu menghabiskan waktu dan sumber daya yang besar

Kinerja yang Baik, Pre-trained model sering kali sudah menguasai banyak pola dan fitur dari dataset besar, sehingga bisa memberikan performa yang baik bahkan pada tugas-tugas baru.