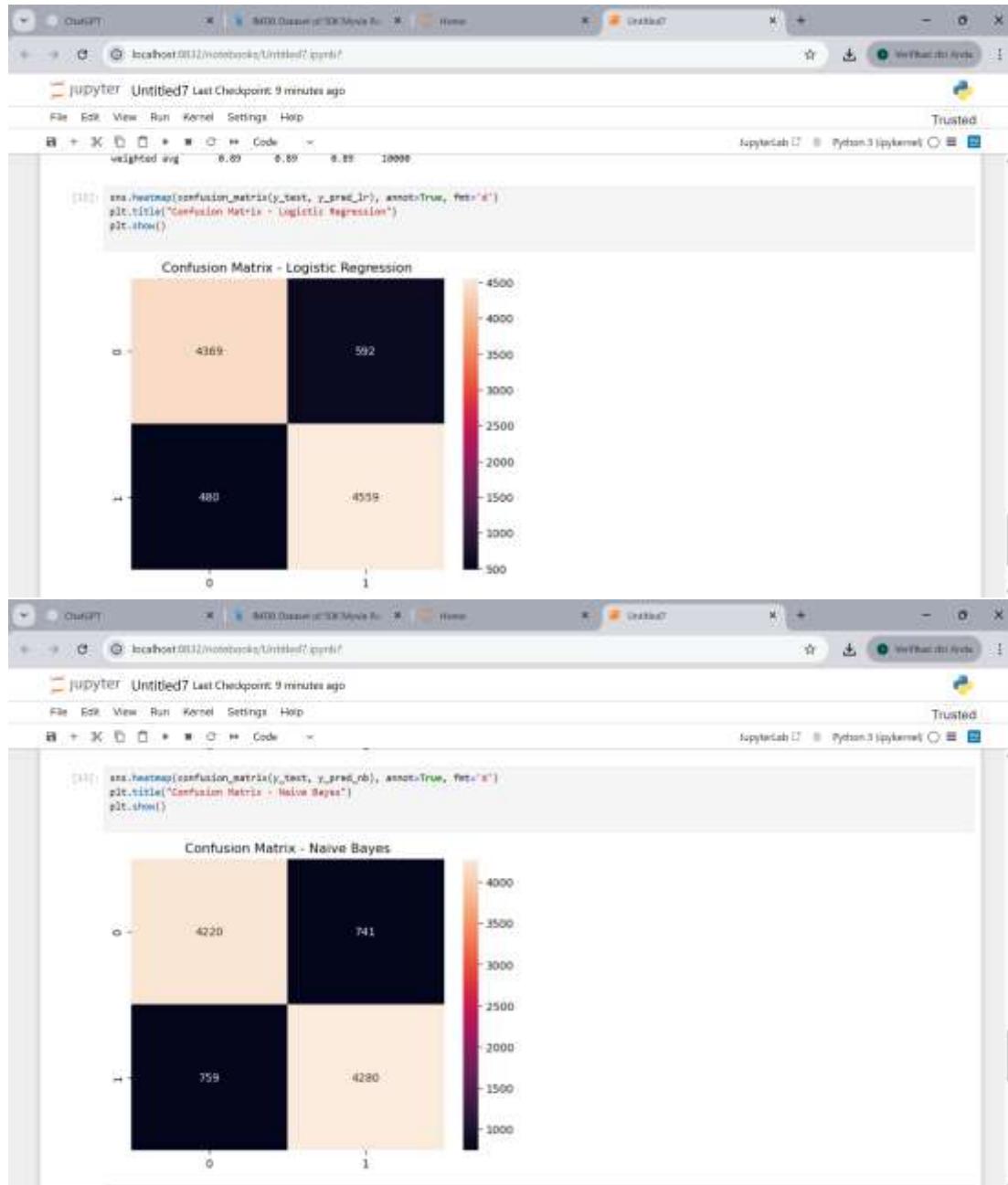


NAMA Riki Raja Purnama

Npm 20123004

ETS bdg



jupyter Untitled7 Last Checkpoint: 9 minutes ago

```
File Edit View Run Kernel Settings Help Trusted
B + X D * M H Code I JupyterLab Python 3 (ipykernel) □
```

```
[1]: sns.heatmap(confusion_matrix(y_test, y_pred_nb), annot=True, fmt='d')
plt.title("Confusion Matrix - Naive Bayes")
plt.show()
```

Confusion Matrix - Naive Bayes

	0	1
0	4220	741
1	759	4280

```
4 Peter Matisse's "Love is the Time of Money" is a... positive peter matisse love is a...
```

```
[2]: tfidf = TfidfVectorizer(max_features=5000)
X = tfidf.fit_transform(df['clean_review'])
y = df['sentiment']

[3]: X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42)

[4]: nb = MultinomialNB()
nb.fit(X_train, y_train)

y_pred_nb = nb.predict(X_test)
print("Naive Bayes vs:")
print(classification_report(y_test, y_pred_nb))

Naive Bayes vs
      precision    recall  f1-score   support
  negative       0.85      0.85      0.85     4961
  positive       0.85      0.85      0.85     5039

  accuracy       0.85      0.85      0.85     10000
  macro avg       0.85      0.85      0.85     10000
```

```

jupyter Untitled7 Last Checkpoint: 8 minutes ago
File Edit View Run Kernel Settings Help
B + X D * M O H Code Trusted
SupportLab Python 3 (ipykernel) 11
segment.

[1]: import re
def clean_text(text):
    text = text.lower()
    text = re.sub(r'[^\w\s]', '', text)
    return text

df['clean_review'] = df['review'].apply(clean_text)
df.head()

[1]:          review sentiment
0 One of the other reviewers has mentioned that... positive one of the other reviewers has mentioned that...
1 A wonderful little production, <br /><br />The... positive a wonderful little production by the fir...
2 I thought this was a wonderful way to spend it... positive I thought this was a wonderful way to spend it...
3 Basically there's a family where a little boy... negative basically there's a family where a little boy ...
4 Peter Matto's 'Lous in the Time of Money' is... positive peter matto's lous in the time of money is a ...

[2]: tfidf = TfidfVectorizer(max_features=5000)
x = tfidf.fit_transform(df['clean_review'])
y = df['sentiment']

[3]: x_train, x_test, y_train, y_test = train_test_split(
    x, y, test_size=0.2, random_state=42)

jupyter Untitled7 Last Checkpoint: 8 minutes ago
File Edit View Run Kernel Settings Help
B + X D * M O H Code Trusted
SupportLab Python 3 (ipykernel) 11

```

A bar chart titled 'sentiment' showing the distribution of reviews. The x-axis has two categories: 'positive' and 'negative'. The y-axis is labeled 'count' and ranges from 0 to 20,000. The 'positive' bar is taller than the 'negative' bar.

sentiment	count
positive	~19,000
negative	~11,000

```

[3]: import re
def clean_text(text):
    text = text.lower()
    text = re.sub(r'[^\w\s]', '', text)
    return text

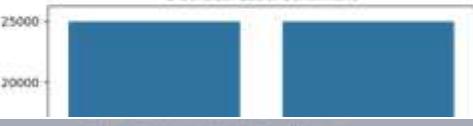
df['clean_review'] = df['review'].apply(clean_text)
df.head()

```

```

jupyter Untitled7 Last Checkpoint: 7 minutes ago
File Edit View Run Kernel Settings Help
B + X D * M H Code <-
In [1]: df.info()
Out[1]:

RangeIndex: 50000 entries, 0 to 49999
Data columns (total 2 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   review    50000 non-null  object 
 1   sentiment  50000 non-null  object 
dtypes: object(2)
memory usage: 781.4+ KB
In [2]: sentiment
Out[2]:
positive    25000
negative    25000
Name: count, dtype: int64
In [3]: sns.countplot(data=df, ax=sentiment)
plt.title("Distribusi Label Sentiment")
plt.show()



```



```

jupyter Untitled7 Last Checkpoint: 7 minutes ago
File Edit View Run Kernel Settings Help
B + X D * M H Code <-
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sn
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix

In [2]: df = pd.read_csv("IMDB Dataset.csv")
df.head()

Out[2]:
review sentiment
0 One of the other reviewers has mentioned that... positive
1 A wonderful little production. <br/><br/>The... positive
2 I thought this was a wonderful way to spend it... positive
3 Basically there's a family where a little boy... negative
4 Peter Mitter's "Love in the Time of Money" is... positive

In [3]: df.info()

```

The image shows a dual-monitor setup. The top monitor displays a Jupyter Notebook interface running on a local host. The bottom monitor displays a Kaggle profile page for a user named Riki raja Purnama.

Jupyter Notebook (Top Monitor):

- Address bar: `localhost:8888/tree?`
- File menu: File, View, Settings, Help.
- Running tab is selected.
- File list:
 - Name
 - anaconda3
 - Cisco Packet Tracer 8.2.2
 - Complete-E-commerce-in-laravel-10
 - Contacts
 - Creative Cloud Files Personal Account (hamidansyah46@gmail.com) C8B91D0E10...
 - Documents
 - Downloads
- Favorites section:
 - fluxion
 - laravel-project
 - laravel-toko-online
 - laravel-toko-online-laravel
 - miniconda3
 - Music
 - nana-projek-anda

Kaggle Profile (Bottom Monitor):

- Profile picture: Riki raja Purnama
- Search bar: Search
- Header: Welcome, Riki raja Purnama!
- Statistics:
 - LOGIN STREAK: 1 day
 - TIER PROGRESS: 0%
 - PUBLIC ACTIVITY: 0
- Summary cards:
 - Datasets: 0 total created
 - Notebooks: 0 out of 100 to get to Expert
 - Competitions: 0 total joined
 - Discussions: 0 total posted
- Sidebar navigation:
 - Create
 - Home
 - Competitions
 - Datasets
 - Models
 - Benchmarks
 - Gamer Arena
 - Code
 - Discussions
 - Learn
 - More
- Work section: How to start: Choose a focus for today.
- Footer: Learn more | OK, I'm fine.

```
FileNotFoundError: [Errno 2] No such file or directory: '/content/WA_Fn-UseC_-Telco-Customer-Churn.csv'

# Import required libraries
import pandas as pd
import numpy as np

# Read CSV file
df = pd.read_csv('/content/WA_Fn-UseC_-Telco-Customer-Churn.csv') # path to dataset file
```