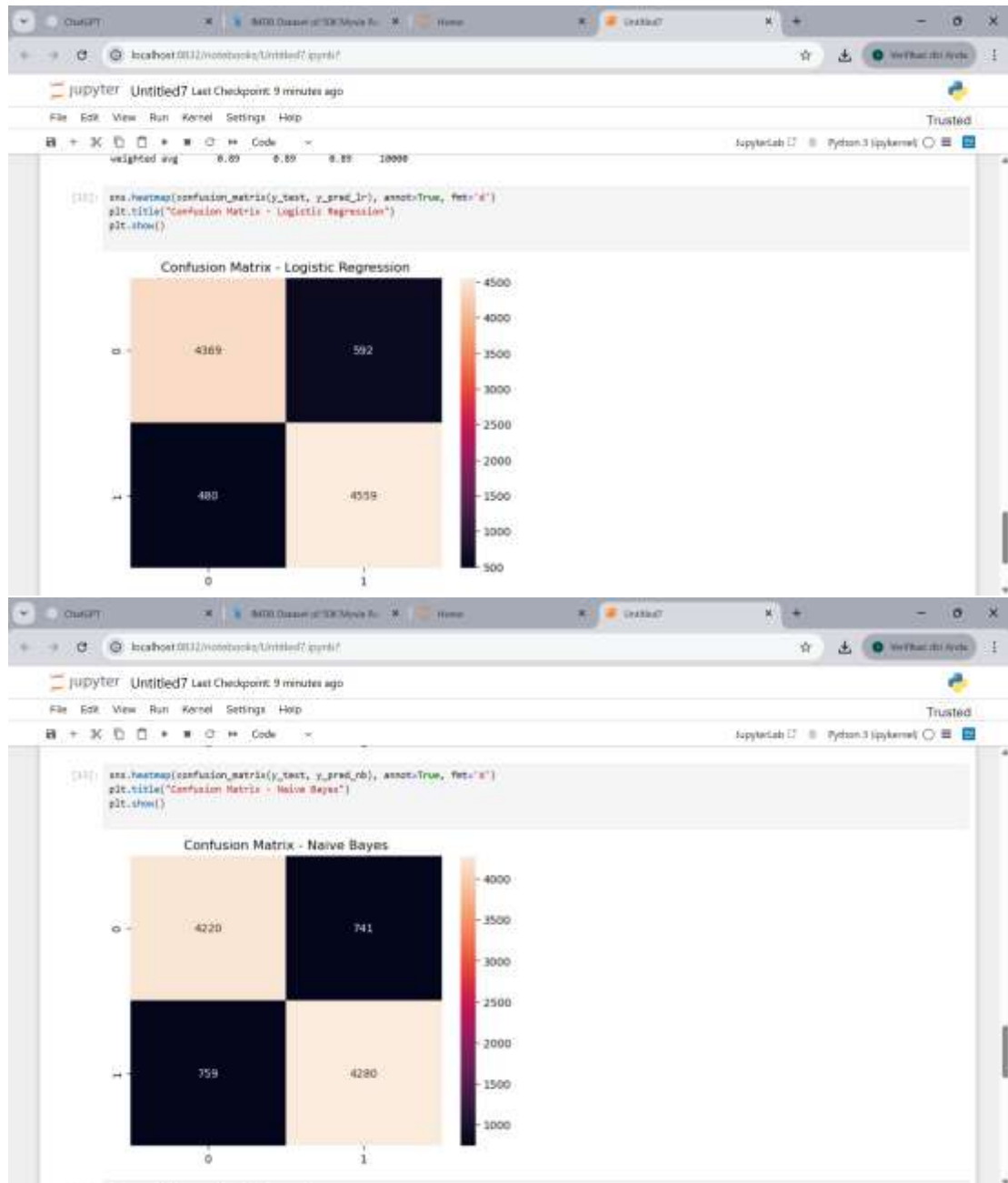
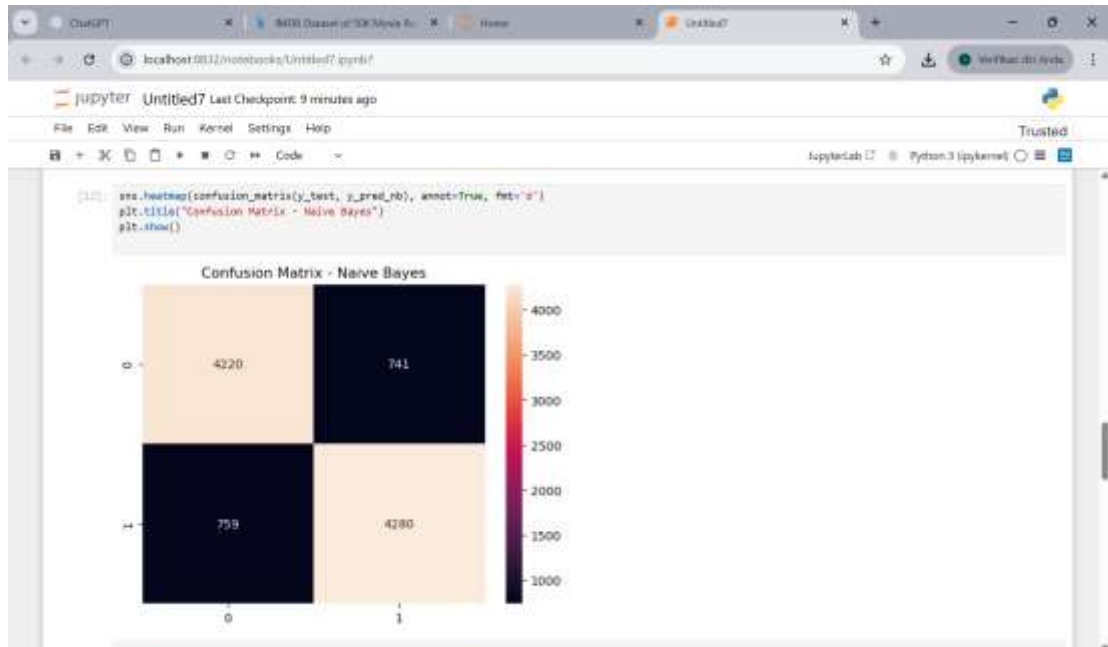


NAMA Riki Raja Purnama

Npm 20123004

ETS bdg





Outlook | 8400 Output of SVM Movie Re... | Home | Untitled7

localhost:8888/notebooks/Untitled7.ipynb

jupyter Untitled7 Last Checkpoint: 8 minutes ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

```
4 Peter Matte's "Love in the Time of Money" is... positive peter matte's love in the time of money is a...

[7]: tfidf = TfidfVectorizer(max_features=10000)
x = tfidf.fit_transform(df['clean_review'])
y = df['sentiment']

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

[9]:

[10]: nb = MultinomialNB()
nb.fit(x_train, y_train)

y_pred_nb = nb.predict(x_test)
print("== Naive Bayes ==")
print(classification_report(y_test, y_pred_nb))

== Naive Bayes ==
precision    recall  f1-score   support

negative    0.85    0.80    0.83    4961
positive    0.85    0.85    0.85    5839

accuracy          0.85    0.85    0.85    10800
macro avg          0.85    0.85    0.85    10800
```

```
import re

def clean_text(text):
    text = text.lower()
    text = re.sub(r'[a-z-25]', '', text)
    return text

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

review sentiment clean_review
0 One of the other reviewers has mentioned that... positive one of the other reviewers has mentioned that...
1 A wonderful little production. <le /><br />The... positive a wonderful little production <br />for the film...
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 Matel's "Love in the Time of Money" is... positive peter matel's love in the time of money is a...

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

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42)
```



