AIML_Exit_Exam.ipynb

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Run all

RAM Disk

Files

- ..
- drive
  - MyDrive
    - AI ML Notes
    - AI Research Papers
    - Colab Notebooks
    - DataSet
    - ICTAK_Project_Data
    - ICT_Project_IntelligentGOSe...
    - Misc
    - Resume
    - exit_exam
      - Reviews.csv
      - Reviews.csv.zip
    - Prototype.pdf
    - Python Quiz Web App
    - Python Quiz Web App for C...
    - Python Quiz Web App for C...
    - Rohit S_Resume (1).pdf

Disk 67.66 GB available

```
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open>=1.8.1->gensim) (1.17.3)
```

[14] ✓ 0s

```
from gensim.models import Word2Vec
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
import numpy as np
```

[16] ✓ 6s

```
# tokenized texts
tokenized_reviews = df['Clean_Text'].apply(lambda x: x.split())
```

[17] ✓ 4m

```
# Train Word2Vec model
w2v_model = Word2Vec(sentences=tokenized_reviews, vector_size=100, window=5, min_count=2, workers=4, seed=42)
```

[18] ✓ 0s

```
# Averaged Word2Vec vector for a review
def get_review_vector(tokens, model, vector_size):
    vectors = [model.wv[word] for word in tokens if word in model.wv]
    if vectors:
        return np.mean(vectors, axis=0)
    else:
        return np.zeros(vector_size)
```

[19] ✓ 1m

```
# Create feature vectors
X_w2v = np.array([get_review_vector(tokens, w2v_model, 100) for tokens in tokenized_reviews])
```

[20] ✓ 0s

```
# Train-test split
X_train_w2v, X_test_w2v, y_train_w2v, y_test_w2v = train_test_split(X_w2v, df['Sentiment'], test_size=0.2, random_state=42)
```

[23]

```
# Random Forest classifier
rf = RandomForestClassifier(n_estimators=20, max_depth=10, n_jobs=-1, random_state=42) #reduced to fix processing speed
rf.fit(X_train_w2v, y_train_w2v)
y_pred_rf = rf.predict(X_test_w2v)
```

[24]

```
# Evaluation
```

Variables Terminal

2:51 PM Python 3

NIFTY +0.74%

Search

ENG IN 14:52 15-10-2025

AIML_Exit_Exam.ipynb

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Run all

RAM Disk

Files

- ..
- drive
  - MyDrive
    - AI ML Notes
    - AI Research Papers
    - Colab Notebooks
    - DataSet
    - ICTAK_Project_Data
    - ICT_Project_IntelligentGOSe...
    - Misc
    - Resume
    - exit_exam
      - Reviews.csv
      - Reviews.csv.zip
    - Prototype.pdf
    - Python Quiz Web App
    - Python Quiz Web App for C...
    - Python Quiz Web App for C...
    - Rohit S_Resume (1).pdf

Disk 67.66 GB available

[18] ✓ 0s

```
        return np.mean(vectors, axis=0)
    else:
        return np.zeros(vector_size)
```

[19] ✓ 1m

```
# Create feature vectors
X_w2v = np.array([get_review_vector(tokens, w2v_model, 100) for tokens in tokenized_reviews])
```

+ Code + Text

[20] ✓ 0s

```
# Train-test split
X_train_w2v, X_test_w2v, y_train_w2v, y_test_w2v = train_test_split(X_w2v, df['Sentiment'], test_size=0.2, random_state=42)
```

[23]

```
# Random Forest classifier
rf = RandomForestClassifier(n_estimators=20, max_depth=10, n_jobs=-1, random_state=42) #reduced to fix processing speed
rf.fit(X_train_w2v, y_train_w2v)
y_pred_rf = rf.predict(X_test_w2v)
```

[24] ✓ 1s

```
# Evaluation
print(classification_report(y_test_w2v, y_pred_rf))
```

```
              precision    recall  f1-score   support

    Negative       0.89      0.33      0.48     16379
    Positive       0.89      0.99      0.94     88784

    accuracy                           0.89    105163
   macro avg       0.89      0.66      0.71    105163
weighted avg       0.89      0.89      0.87    105163
```

Key advantage of using Word2Vec embeddings over TF-IDFis that it captures the semantic meaning and relationships between words. and it is done through mapping them to continuous vector spaces based on their context in real text. Whereas TF-IDF considers only frequency and not semantic relationship.

[ ]

Variables Terminal

2:51 PM Python 3

Gold +0.62%

Search

ENG IN 14:52 15-10-2025