Московский государственный технический университет им. Н.Э. Баумана
Кафедра «Системы обработки информации и управления»

Рубежный контроль №2
по дисциплине
«Методы машинного обучения»
на тему

«Методы обработки текстов»

Выполнил:
студент группы ИУ5и-22М
Джин Шуо

Москва — 2024 г.

Варианты заданий

Группа	Классификатор№1	Классификатор№2
ИУ5И-22М	RandomForestClassifier	LogisticRegression

текст программы

```
from sklearn.datasets import fetch_20newsgroups

from sklearn.model_selection import train_test_split

from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer

from sklearn.ensemble import RandomForestClassifier

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import classification_report, accuracy_score


categories = ['alt.atheism', 'comp.graphics', 'sci.space', 'talk.religion.misc']

newsgroups = fetch_20newsgroups(subset='all', categories=categories)

X_train, X_test, y_train, y_test = train_test_split(newsgroups.data, newsgroups.target,
test_size=0.3, random_state=42)


# CountVectorizer

count_vectorizer = CountVectorizer()

X_train_counts = count_vectorizer.fit_transform(X_train)

X_test_counts = count_vectorizer.transform(X_test)


# TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()

X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)

X_test_tfidf = tfidf_vectorizer.transform(X_test)


# CountVectorizer

# Классификатор№1RandomForestClassifier

rf_classifier = RandomForestClassifier(random_state=42)

rf_classifier.fit(X_train_counts, y_train)
```

```
y_pred_rf_counts = rf_classifier.predict(X_test_counts)
print("RandomForestClassifier with CountVectorizer")
print(classification_report(y_test, y_pred_rf_counts))
print("Accuracy:", accuracy_score(y_test, y_pred_rf_counts))
# Классификатор№2LogisticRegression
lr_classifier = LogisticRegression(max_iter=1000, random_state=42)
lr_classifier.fit(X_train_counts, y_train)
y_pred_lr_counts = lr_classifier.predict(X_test_counts)
print("LogisticRegression with CountVectorizer")
print(classification_report(y_test, y_pred_lr_counts))
print("Accuracy:", accuracy_score(y_test, y_pred_lr_counts))

# TfidfVectorizer
# Классификатор№1RandomForestClassifier
rf_classifier.fit(X_train_tfidf, y_train)
y_pred_rf_tfidf = rf_classifier.predict(X_test_tfidf)
print("RandomForestClassifier with TfidfVectorizer")
print(classification_report(y_test, y_pred_rf_tfidf))
print("Accuracy:", accuracy_score(y_test, y_pred_rf_tfidf))
# Классификатор№2LogisticRegression
lr_classifier.fit(X_train_tfidf, y_train)
y_pred_lr_tfidf = lr_classifier.predict(X_test_tfidf)
print("LogisticRegression with TfidfVectorizer")
print(classification_report(y_test, y_pred_lr_tfidf))
print("Accuracy:", accuracy_score(y_test, y_pred_lr_tfidf))
```

экранные формы с примерами выполнения программы

```
RandomForestClassifier with CountVectorizer
              precision    recall  f1-score   support

           0       0.88      0.92      0.90       224
           1       0.82      0.98      0.89       297
           2       0.95      0.90      0.93       307
           3       0.93      0.68      0.78       189

    accuracy                           0.89      1017
   macro avg       0.90      0.87      0.88      1017
weighted avg       0.89      0.89      0.88      1017

Accuracy: 0.8869223205506391
LogisticRegression with CountVectorizer
              precision    recall  f1-score   support

           0       0.91      0.94      0.92       224
           1       0.96      0.98      0.97       297
           2       0.96      0.95      0.95       307
           3       0.91      0.86      0.88       189

    accuracy                           0.94      1017
   macro avg       0.93      0.93      0.93      1017
weighted avg       0.94      0.94      0.94      1017
...
   macro avg       0.93      0.92      0.92      1017
weighted avg       0.93      0.93      0.93      1017

Accuracy: 0.9311701081612586
```

Список литературы

[1] Гапанюк Ю. Е. LAB_MMO__DATA_STORYЛабораторная работа №1Создание "истории о данных" (Data Storytelling)// GitHub. — 2024. — Режим доступа:https://github.com/ugapanyuk/courses_current/wiki/MMO_RK_2