

**[Day-40 2211cs020196]Write a Python program to classify text using an ensemble model ( RandomForestClassifier ) from sklearn.ensemble. Use TfidfVectorizer to convert text into numerical format.Example input: ["Buy now and win a prize", "Normal email content", "Congratulations, you won!"]** ¶

```
In [1]: 1 from sklearn.feature_extraction.text import TfidfVectorizer
2 from sklearn.ensemble import RandomForestClassifier
3 from sklearn.model_selection import train_test_split
4 documents = ["Buy now and win a prize", "Normal email content", "Congratulations, you won!",
5             "Limited time offer", "Just a regular message", "You have been selected for a reward"]
6 labels = ["spam", "not_spam", "spam", "spam", "not_spam", "spam"]
7 vectorizer = TfidfVectorizer()
8 X = vectorizer.fit_transform(documents)
9 X_train, X_test, y_train, y_test = train_test_split(X, labels, test_size=0.2, random_state=42)
10 classifier = RandomForestClassifier()
11 classifier.fit(X_train, y_train)
12 new_text = ["Exclusive offer just for you"]
13 X_new = vectorizer.transform(new_text)
14 predicted_category = classifier.predict(X_new)
15 print(f"The predicted category for the new text '{new_text[0]}' is: {predicted_category[0]}")
16
```

The predicted category for the new text 'Exclusive offer just for you' is: spam

In [ ]:

1