In [8]:
```
import numpy as np
import nltk
from sklearn.datasets import load_files
nltk.download('stopwords')
from nltk.corpus import stopwords
nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Monipraba\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\Monipraba\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[8]:
```
True
```

In [9]:
```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.ensemble import RandomForestRegressor
```

In [10]:
```
df=pd.read_csv(r"C:\Users\Monipraba\Documents\projectdataset.csv")
answers = df.head(30)
answers
```

Out[10]:

	S.NO	QUESTION	ANSWER	MARKS	Unnamed: 4
0	1.0	What two motions do all planets have?	All planets have two types of motion, known as...	2.0	NaN
1	2.0	When did the Space Age begin?	The space age began on October 4, 1957, when t...	5.0	NaN
2	3.0	What is the visible part of the Sun called?	The outer region of the Sun that is normally v...	1.0	NaN
3	4.0	What makes a planet a dwarf planet?	In 2006, Pluto, Eris, and Ceres were classifie...	3.0	NaN
4	5.0	What term describes the alignment of three cel...	When three celestial bodies appear to be in a ...	4.0	NaN
5	6.0	Which of these objects is the farthest from t...	90377 Sedna lies more than 11 billion kilomete...	4.0	NaN
6	7.0	Approximately how many miles are there in a li...	A light-year is the distance light travels in ...	2.0	NaN
7	8.0	Which is the name of a radio source that is ve...	A quasar is a radio source that comes to Earth...	2.0	NaN
8	9.0	The day on which the Sun?s direct rays cross t...	On the equinox, the day on which the Sun?s pat...	3.0	NaN
9	10.0	Who invented the telescope?	Hans Lippershey (c. 1570-c. 1619) was a specta...	1.0	NaN
10	11.0	what is the Jurassic period named?	The Jurassic period, of Jurassic Park fame, is...	5.0	NaN
11	12.0	Who invented the World Wide Web?	In the early 1990s, a group of computer scient...	5.0	NaN
12	13.0	Which technological developments came first?	The first telescope is thought to have been bu...	5.0	NaN
13	14.0	Who invented the geodesic dome?	R. Buckminster Fuller, an American architect a...	5.0	NaN
14	15.0	What airplane has not been flown commercially ...	The first commercial jet to travel faster than...	4.0	NaN
15	16.0	Moths are a member of what order?	Moths are of the same order as butterflies, th...	3.0	NaN
16	17.0	When was the first plastic made of artificial ...	In 1909 a chemist named Leo H. Baekeland devel...	3.0	NaN
17	18.0	How many litres of milk drinks baby blue whale...	A baby blue whale drinks approximately 190 lit...	3.0	NaN
18	19.0	whom is the centigrade system of temperature m...	Anders Celsius, a Swedish scientist, conceived...	2.0	NaN
19	20.0	Which scientist was born the year Galileo died?	In the year Galileo died, 1642, there was born...	2.0	NaN
20	21.0	For whom is the Fahrenheit system of temperatu...	Daniel Fahrenheit, a German physicist, propose...	5.0	NaN
21	22.0	Who is considered the ?father? of the scienti...	Galileo is considered the father of the experi...	2.0	NaN
22	23.0	In which field did Marie Curie and her daughte...	Marie Curie, also known as Madame Curie, and h...	3.0	NaN
23	24.0	Who developed the theory of evolution?	Charles Darwin (1809?1882) developed the theor...	3.0	NaN
24	25.0	Who invented the safety elevator?	The safety elevator, which will not crash even...	2.0	NaN
25	26.0	Which scientist is well known for his work wit...	Galileo discovered the natural laws that gover...	5.0	NaN

	S.NO	QUESTION	ANSWER	MARKS	Unnamed: 4
26	27.0	Who said, "God does not play dice with the uni...	Albert Einstein, the eminent physicist, said, ...	3.0	NaN
27	28.0	What is the reflectivity of Earth's surface c...	The albedo effect refers to the reflectivity o...	3.0	NaN
28	29.0	What is diatomaceous earth made of?	Diatomaceous earth is made from the fossilized...	5.0	NaN
29	30.0	What is Earth?s predominant environment?	Most of the world is covered by water, and mos...	5.0	NaN

In [11]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
texts = ["When three celestial bodies appear to be in a straight line, as sometimes happens, this is called a syzygy."
         "When three celestial appear to be in normal line as it happens is known as a syzygy."]
tfidf = TfidfVectorizer()
features = tfidf.fit_transform(texts)
pd.DataFrame(features.todense(),columns=tfidf.get_feature_names())
```

Out[11]:

	appear	as	be	bodies	called	celestial	happens	in	is	it	known	line	normal	sometimes	st
0	0.213801	0.213801	0.213801	0.30049	0.30049	0.213801	0.213801	0.213801	0.213801	0.000000	0.000000	0.213801	0.000000	0.30049	0.3
1	0.218603	0.437207	0.218603	0.00000	0.00000	0.218603	0.218603	0.218603	0.218603	0.307239	0.307239	0.218603	0.307239	0.00000	0.(

In [12]:
```
from sklearn.feature_extraction.text import CountVectorizer
texts = ["When three celestial bodies appear to be in a straight line, as sometimes happens, this is called a syzygy."
         "When three celestial appear to be in normal line as it happens is known as a syzygy."]
vectorizer = CountVectorizer()
counts = vectorizer.fit_transform(texts)
pd.DataFrame(counts.todense(),columns=vectorizer.get_feature_names())
```

Out[12]:

	appear	as	be	bodies	called	celestial	happens	in	is	it	known	line	normal	sometimes	straight	syzygy	this	three	to	when
0	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1
1	1	2	1	0	0	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1

In [17]:
```
import spacy
from spacy import displacy

NLP = spacy.load("en_core_web_sm")
doc = NLP("When three celestial appear to be in normal line as it happens is known as a syzygy.")
displacy.serve(doc, style="dep")
```

displaCy
When SCONJ three NUM celestial NOUN appear VERB to PART be AUX in ADP normal ADJ line NOUN as SCONJ it PRON happens VERB is AUX known VERB as ADP a DET syzygy. NOUN advmod nummod nsubj advcl aux xcomp prep amod pobj mark nsubj advcl auxpass prep det pobj

```
Using the 'dep' visualizer
Serving on http://0.0.0.0:5000 (http://0.0.0.0:5000) ...

Shutting down server on port 5000.
```

In [15]:
```
X = answers['ANSWER']
y = answers['MARKS']

# convert text data into numerical features
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(X)

# train the random forest regressor
rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(X, y)

# evaluate the model performance
score = rf.score(X, y)
print('Model Score:', score)

# use the trained model to predict grades for new answers
new_answer = 'The quick brown fox jumps over the lazy dog'
new_answer_vectorized = vectorizer.transform([new_answer])
new_grade = rf.predict(new_answer_vectorized)[0]
print(new_grade)

# assign grades based on predicted scores
if new_grade >= 90:
    grade = 'A'
elif new_grade >= 80:
    grade = 'B'
elif new_grade >= 70:
    grade = 'C'
elif new_grade >= 60:
    grade = 'D'
else:
    grade = 'F'

print('Predicted Grade:', grade)
```

```
Model Score: 0.8346250000000001
3.22
Predicted Grade: F
```

In [ ]: