


Проект по определению времени английского предложения

Подготовили: Глизница М.Н., Снегирева Д.А.

Группа: КЭ-120

Предметная область

 Проверочная работа
ВАРИАНТ 2

Задание 1. Выберите и обведите правильный вариант.

1. They have got a lot of
a) horse **b) toy** c) corn
2. Have they got ... oranges?
a) many b) much **c) an**
3. Have they got ... orange juice?
a) many b) much c) a
4. She has got ... doll.
a) many b) much **c) a**
5. He must ... English.
a) speaks **b) speak** c) to speak
6. I ... to sing the song.
a) would like b) can c) must
7. I must not ... to the park.
a) go b) goes c) to go
8. The boys ... strong.
a) have b) is c) are
9. He ... hockey.
a) play b) to play c) plays
10. He would like ... hockey.
a) play b) to play c) plays

6 11

Итоговая контрольная работа по английскому языку для 7 класса
Учебник: Spotlight-7

A. Listen to Erica talking to her friend in a music shop. What type of music do her friends like?

- | | |
|--------------|-----------------------|
| 1. Paul | a. classical |
| 2. Susan | b. pop music |
| 3. Pat | c. jazz |
| 4. Stephanie | d. rap |
| 5. Alan | e. heavy metal music. |

V1

B. Put the verbs in brackets into the correct tense.

6. He usually(drink) coffee in mornings.
7. At the moment they..... (play) football.
8. Jane(read) the newspaper when the phone rang.
9. We(move) house in 2000.
10. Theyjust(plant) flowers.

C. Choose the right option.

11. Brian is a very boy and the best student in his class.
a. intelligent b. glance c. pipe
12. I like beautiful..... in our village.
a. public b. dictionary c. landscapes
13. John lives in the country because he likes the friendly and people.
a. helpful b. heavy c. crowded
14. He can speak three: English, German and French.
a. languages b. knowledge c. nationality
15. The athlete broke the world for the 100 m race.
a. idea b. record c. advice
16. There's an with Brad Pitt in today's newspaper.
a. interview b. message c. weather
17. Swimming is not in this river.
a. allowed b. polluted c. polite
18. Famous people such as actors and sportsmen are for teenagers.
a. role models b. school assignments c. headphones
19. ... they travel by bicycle every weekend.
a. To keep fit b. Figure skating c. Hardly sleep
20. I find the most stressful.
a. taking care with your parents b. getting enough sleep c. arguing with parents
21. The book is very

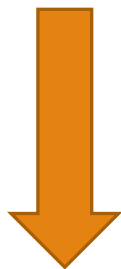
Набор данных



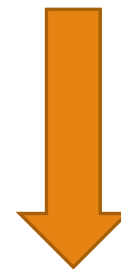
Tatoes

Предобработка данных

Index	id	lang	Text
0	1277	eng	I have to go to sleep.



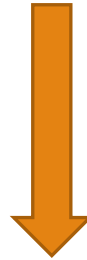
Index	Id	Tag
0	1277	present simple



Index	Text	Tag
0	I have to go to sleep.	present simple

Обработка данных

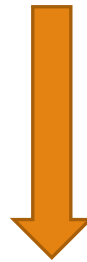
Despite me watching TV, my dog does not bite.



```
[ 'IN', 'PRP', 'VBG', 'NN', ',', 'PRP$',  
  'NN', 'VBZ', 'RB', 'VB', '. ' ]
```

Использование токенайзера

```
['IN', 'PRP', 'VBG', 'NN', ',', 'PRP$',  
 'NN', 'VBZ', 'RB', 'VB', '.']
```



```
array([[0., 1., 1., 1., 0., 0., 1., 1., 1.,  
 1., 0., 0., 1., 0., 0., 1., 0., 0., 0., 0.,  
 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,  
 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,  
 0., 0., 0., 0., 0., 0., 0., 0.]])
```

Создание и обучение модели

```
XGBClassifier(  
    random_state = 42,  
    seed = 2,  
    colsample_bytree = 0.6,  
    subsample = 0.7,  
    verbosity = 0  
    n_estimators = 300  
)
```

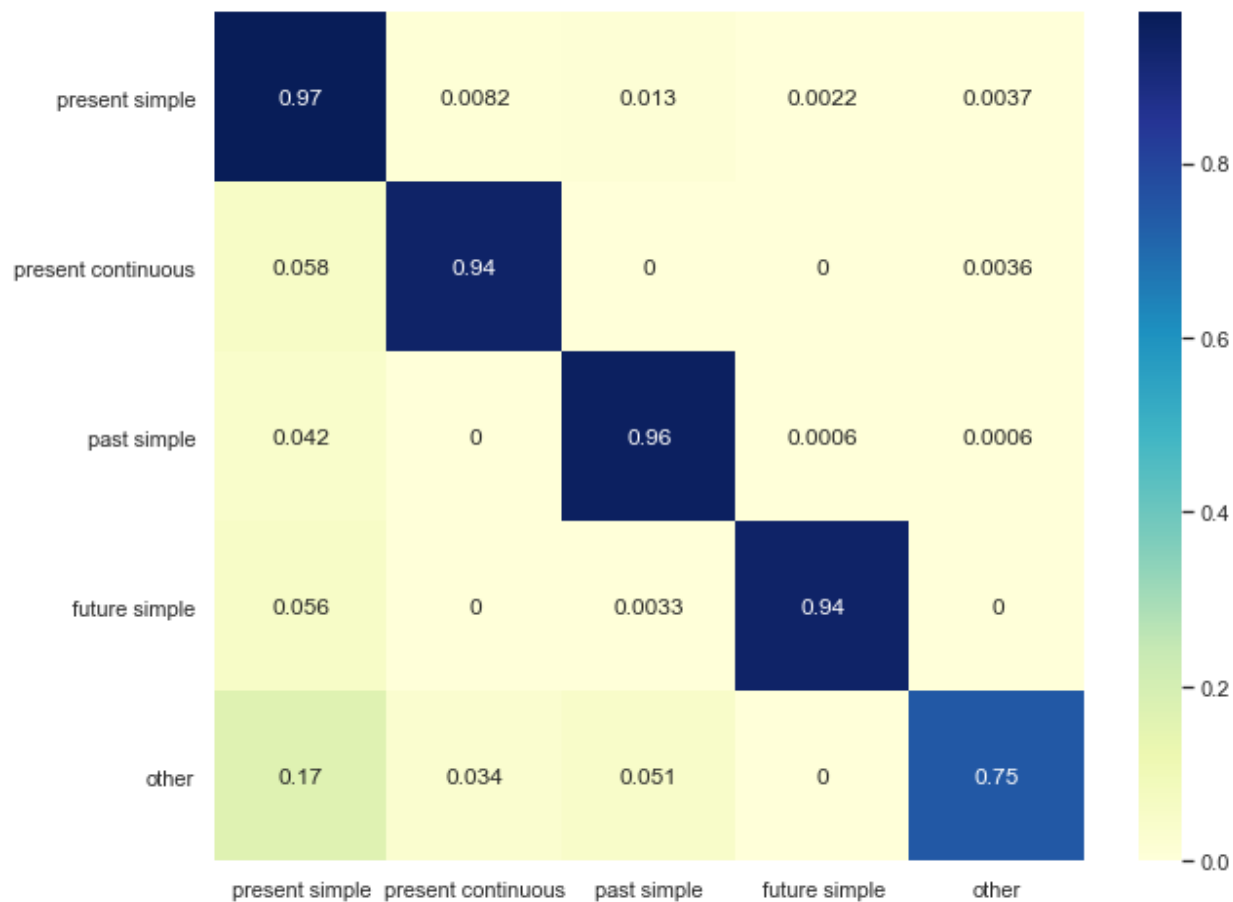
Количественные характеристики работы решения

	precision	recall	f1-score	support
present simple	0.97	0.97	0.97	6705
present continuous	0.96	0.94	0.95	1648
past simple	0.95	0.96	0.95	1677
future simple	0.95	0.94	0.94	306
other	0.58	0.75	0.65	59
accuracy			0.96	10395

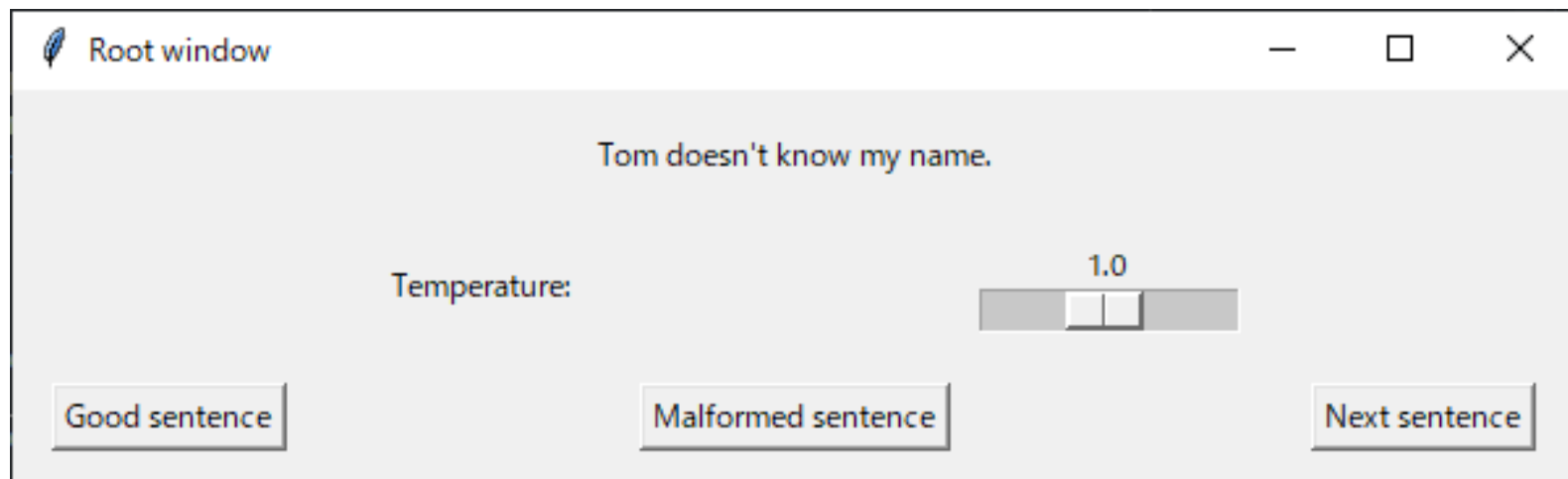
Время обучения: 72 сек.

Время классификации одного предложения: меньше 1 сек.

Матрица ошибок



Генератор на основе полученного классификатора



Использованные инструменты и библиотеки

1. Python 3.7.9

2. Pandas

3. nltk

4. Keras

5. numpy

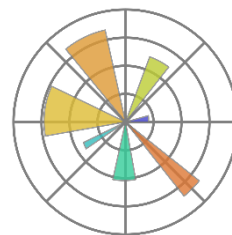
6. matplotlib

7. sklearn

8. xgboost



Keras



pandas

Итоговая трудоемкость

Всего было затрачено 16 человеко-часов, написано 245 строк кода.

Работа была распределена следующим образом:

- обработка данных – Глизница М. Н.;
- создание и обучение модели – Снегирева Д. А.;
- доработка кода – оба участника команды.

Заключение

В результате проектной работы был создан классификатор для определения времени английского предложения.

В дальнейшем классификатор был использован при разработке генератора предложений во времени Present Simple.

Литература

1. Murphy R., English Grammar in Use: A self-study reference and practice book for intermediate students : with answers, 5th ed. Cambridge: Cambridge University Press, 2019.
2. Bird S., Loper E., Klein E. Natural Language ToolKit (NLTK) Book / S. Bird, E. Loper, E. Klein, 2009.
3. Tatoeba [Электронный ресурс]. URL: www.tatoeba.org (дата обращения: 04.05.2021).
4. XGBoost: A Scalable Tree Boosting System [Электронный ресурс]. URL: <https://arxiv.org/abs/1603.02754> (дата обращения: 27.05.2021).