

REPORT OF THE CLASSIFICATION OF A NİTEL-NİCEL DATASET USING BERTURK MODEL STUDY

Murat Can Zöhre

2019555458

INDEX

1. Dataset	3
2. Data Cleaning	3
2.1.Data Cleaning Function	3
2.2.Remove Stopwrods Function	4
2.3.Remove Punctuation Function	4
3. Approaches and Operations	5
3.1.Approaches	5
3.1.1. Two Label Approach	5
3.1.1.1. Single-Column Approach	6
3.1.1.2. Multi-Column Approach	7
3.2.Operations	7
3.2.1. Functions	7
3.2.1.1. Lemmatizer Function	7
3.2.1.2. Keyphrase Extraction Function	8
3.2.2. List Of Operations	9
4. Results	10
4.1.Results Of A Five Label Approach	10
4.1.1. Results Of A Single-Column Approach	10
4.1.2. Results Of A Multi-Column Approach	12
4.2.Results Of A Two Label Approach	13
4.2.1. Results Of A Single-Column Approach	13
4.2.2. Results Of A Multi-Column Approach	14

1. DATASET

Nitel-Nicel Dataset has used for this study.This dataset has 5 columns.

Example row of a dataset:

	Nitel_Soru	Nitel_Cevap	Nicel_Soru	Nicel_Cevap_KaanAla	Nicel_Puan
0	Bu göreve gelirken ve görevinizi yürütürken çe...	Kadınların önünde örnek yok; teyzemizden halam...	Kadınlara rol modeli olabilecek yeterli sayıda...	(1) Hiç Katılmıyorum	1

Nitel_Soru: Column representing the qualitative question asked

Nitel_Cevap: Column representing the qualitative answer

Nicel_Soru: Column representing the quantative question

Nicel_Cevap_KaanAla : Column representing quantative answer

Nicel_Puan: Column representing the numeric representation of a Nicel_Cevap_KaanAla

In this study first 3 columns used as an input and Nicel_Puan used as a label.

2. DATA CLEANING

As a starting step, data that distorts the context of the input or does not contribute to it are removed from the dataset. Such as punctuation marks, irrelevant numbers, and some symbols like('>' , '<' etc.). Irregularity in the shape of the inputs has also been eliminated.To implement cleaning operations various functions has applied to the dataset.

2.1 DATA CLEANING FUNCTION

With the help of this function the irrelevant numbers in the data were cleaned and the irregularities in the entries were fixed

```
: def data_cleaning(x):  
    x=''.join((s for s in x if not s.isdigit()))  
    x=x.replace(".", "",1)  
    x = x.replace("<", "")  
    x=x.strip('-')  
    x=x.strip()  
    return x
```

```
: df["Nitel_Soru"] = df['Nitel_Soru'].apply(data_cleaning)  
df["Nitel_Cevap"] = df['Nitel_Cevap'].apply(data_cleaning)  
df["Nicel_Soru"] = df['Nicel_Soru'].apply(data_cleaning)  
df
```

2.2 REMOVE STOPWORDS FUNCTION

Stopwords are removed from the inputs data with using remove_stopwords function. Since stopwords do not have any contextual contribution to the input data they were removed to simplify the input data and increase efficiency of the model.

```
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize, sent_tokenize
nltk.download('stopwords')
from nltk.corpus import stopwords
stopwords = set(stopwords.words('turkish'))
def remove_stopwords(text):
    words = word_tokenize(text)
    filtered_text=[]
    for word in words:
        if word not in stopwords:
            filtered_text.append(word)
    text=" ".join(filtered_text)
    return text
```

```
df["Nitel_Soru"] = df['Nitel_Soru'].apply(remove_stopwords)
df["Nitel_Cevap"] = df['Nitel_Cevap'].apply(remove_stopwords)
df["Nitel_Soru"] = df['Nitel_Soru'].apply(remove_stopwords)
```

2.3 REMOVE PUNCTUATION FUNCTION

All punctuation marks were discarded from the input data for simplicity. Also with the use of this function data cleaning function were double checked.

```
punct_list = list(string.punctuation)
def remove_punct(text):
    words=word_tokenize(text)
    text_filtered=[]
    for word in words:
        if word not in punct_list:
            text_filtered.append(word)
    text=" ".join(text_filtered)
    return text
```

```
df["Nitel_Soru"] = df['Nitel_Soru'].apply(remove_punct)
df["Nitel_Cevap"] = df['Nitel_Cevap'].apply(remove_punct)
df["Nitel_Soru"] = df['Nitel_Soru'].apply(remove_punct)
```

After data cleaning operations were applied the data which is going to be used as a base data is ready. But since the aim of this study was to reach the optimum accuracy value, various approaches were applied to the data through additional functions.

3. APPROACHES and OPERATIONS

Since the aim of this study was to reach the optimum accuracy value ; multiple , distinct results to compare required.To produce distinct results , distinct inputs are required and to produce distinct inputs , distinct operations and approaches are required.

Distinction of approach and operation is very significant. Approach means how to handle problem , it implies a perspective but operation means content of a certain approaches which can be a combination of functions or just one one function or various transactions.

In this study the results to be compared were produced as the result of each operation so each result corresponds to a particular operation.

3.1 APPROACHES

There are two main approach in this study and two other approaches under main approaches and each sub approaches includes operations.

The original dataset has 5 labels as a target value which could increase complexity. In order to observe this problem, five target labels were reduced to two target labels.

This two approaches are the main approaches of the study , 2 label approach and 5 label approach.Each operation was performed separately under these two approaches.

3.1.1 TWO LABEL APPROACH:

In order to apply 2 label approach a function which reduce the target values from five distinct values to two distinct values was required. The original target values are:

- 0: Hiç Katılmıyorum
- 1: Katılmıyorum
- 2: Kararsız
- 3: Katılıyorum
- 4: Tamamen Katılıyorum

To reduce these 5 distinct value to 2 distinct value first the rows with a label value are 2 removed.Then label values 0 and 1 combined as a distinct label and 3 and 4 combined as a other distinct label.So label 0 and label 1 become label 0 as Katılmıyorum and label 3 and label 4 become label 1 as Katılıyorum.The target values of a 2 label approach are:

- 0: Katılmıyorum
- 1: Katılıyourm

```
df= df[df['labels'] != 2]
def label_reduction(label):
    if label==0 or label==1:
        label=0
    else:
        label=1
    return label
df["labels"] = df["labels"].apply(label_reduction)|
```

In addition to two main approaches in this study , there are two other approaches has used as a subset of the main approaches.

Since input data consist of three columns there could be two approaches to handle input data which has multiple columns.

First approach is single column approach.In this approach columns are merged into a single column and treat the input data as a single column.

Second approach is multi column approach.In this approach columns handled separately and push into a model as distinct inputs.

These two approaches were applied separately under each of the main approaches.

3.1.1.1 SINGLE-COLUMN APPROACH

In single column approach , three input columns which are Nitel_Soru , Nitel_Cevap and Nicel_Soru merged into a single column called as inputs and all operations performed on this inputs column.

Nitel_Soru	Nitel_Cevap	Nicel_Soru	labels	input
0	Bu göreve gelirken görevinizi yürütürken çevre...	Kadınların önünde örnek yok teyzemizden halamı...	Kadınlara rol modeli olabilecek yeterli sayıda...	0
1	Sektörde sektördeki kadın çalışan oranı belirt...	Nihayetinde yönetsel branşlarda fazla kadın ol...	Kadınların okul yöneticisi olarak yeterli oran...	0
2	Sektörde sektördeki kadın çalışan oranı belirt...	Bizde zaten kadın çalışan oranı farklı değil h...	Kadınların okul yöneticisi olarak yeterli oran...	0
3	Kadın yönetici olarak özellikle cinsiyet etmen...	Bir yandan erkek egemen şirketlerde zaman zama...	Kadınlar okul yöneticisi olmak pozitif ayrımcı...	0
4	Kadın yönetici olarak özellikle cinsiyet etmen...	Ekelere verilen imkânlar hiçbir zaman kadına ...	Kadınlar okul yöneticisi olmak pozitif ayrımcı...	0
...
560	Başarılarınızda başarısızlıklarınızda diğer in...	...kimse üzgün değildir...	Başarıların başarısızlıkların üzerinde diğer i...	4
561	Başarılarınızda başarısızlıklarınızda diğer in...	...aslında zor felsefi bir soru Akademik çalışma...	Başarıların başarısızlıkların üzerinde diğer i...	4
562	Yaptığınız işin sonuçları üzerinde oynadığınız...	yapmış olduğun işlerin sonuçlarında kendi ko...	Yaptığım işlerin sonuçları üzerinde kontrolümü...	4
563	Başarılarınızda şansın etkisi nedir	...fırsatları iyi görebilmeyi şans olarak görüyö...	Başarıların çoğu şansla ilişkilidir	4
564	Başarılarınızda şansın etkisi nedir	...hayatımın seyrini değiştirecek insanlarla tan...	Başarıların çoğu şansla ilişkilidir	4

```
df["input"]=df["Nitel_Soru"]+" "+df["Nitel_Cevap"]+" "+df["Nicel_Soru"]
df=df.drop(["Nitel_Soru","Nitel_Cevap","Nicel_Soru"],axis=1)
df
```

3.1.1.2 MULTI-COLUMN APPROACH

In multiple columns approach each column is treated as a different input but pushed together to the model. Main purpose of this approach is to preserve context of each column.

To implement multiple column approach, BERT models special [SEP] token was used. [SEP] token ensures that each column is handled individually.

When using [SEP] token, columns are merged as in single column approach with only one difference; when merging columns, [SEP] token is placed between columns. Through this feature BERT model understands that inputs are independent from each other when it sees [SEP] token.

```
df["input"]="[CLS]"+" "+df["Nitel_Soru"]+" "+"[SEP]"+" "+df["Nitel_Cevap"]+" "+"[SEP]"+" "+df["Nice1_Soru"]+" "+"[SEP]"
|labels,input
0,[CLS] Bu göreve gelirken görevinizi yürütürken çevrenizden destek aldınız alıyorsunuz [SEP] Kadınların önünde örnek yok teyzemizden halamızdan böyle şeyleri görmüyoruz İş hayatı
yönlendirebiliriz yetiştirebiliriz böyle bir yönlendirme olmuyor [SEP] Kadınlara rol modeli olabilecek yeterli sayıda kadın okul yöneticisi vardır [SEP]
0,[CLS] Sektörde sektördeki kadın çalışan oranı belirtilecek kadın çalışan varken sektörde kadın yönetici oranı belirtilecek oranında kadın yönetici mevcut Sizce kadınların yönetim
kademelerinde temsil edilmelerinin sebepleri neler olabilir [SEP] Nihayetinde yönetsel branşlarda fazla kadın olmaması -mühendislik branşlarında- konuda yetmiş bayanın olması yönetim
kademelerine yansıyor düşünüyorum [SEP] Kadınların okul yöneticisi olarak yeterli oranda temsil edildiğini düşünüyorum [SEP]
```

3.2 OPERATIONS

Under each approach there are various operations which produce results for the study. Operations in this study include two main functions; Lemmatizer function and keyphrase extraction function. In addition to the operations where the data is used as it is, the variety of results is increased by operations which include these functions separately and together to the data.

3.2.1 FUNCTIONS

3.2.1.1 LEMMATIZER FUNCTION

To push input data into a model, data needs to be converted into numerical values. To implement this conversion there is a tokenizer function which converts each token in the data into a numerical value.

Tokenizer function of a BertTürk model assigns different numerical values to words with the same lemma but different suffixes that do not affect the context of the input. This situation could, dysfunctionally increase complexity of an input data and may adversely affect the outcome of a model. To overcome this problem lemmatizer function was used. Lemmatizer function lemmatizes each word in the input data and equalizes the words with the same roots numerically.

Definition of a lemmatizer function:

```
import zeyrek
analyzer = zeyrek.MorphAnalyzer()
def lemmatizer(text):
    words_lemmatize=[]
    words = word_tokenize(text)
    for word in words:
        word=analyzer.analyze(word)[0][0].lemma
        words_lemmatize.append(word)
    text=" ".join(words_lemmatize)
    return text
```

3.2.1.2 KEYPHRASE EXTRACTION FUNCTION

One of the possible problem about the original data is original data consists of a lots of word so input vector has too many dimension.This situation could increase complexity of a input data and may adversely effects the model performance.To overcome this possible problem keyphrase extraction function was used.

With using keyphrase extraction function the size of the input is reduced while preserving its essence and context

When using keyphrase extraction function under single colum approach there are two possible ways ; first way is applying keyphrase extraction function to a three columns seperately then merging columns , the second way is applying keyphrase extraction to a input column after merging columns.In multiple column approach since the way of applying keyphrase extraction after merging columns is against the purpose of approach just first way was used under this approach.

Applying keyphrase extraction function after merge columns reduce the dimension of the input data more than applying keyphrase extraction functions to three columns separately then merging.This situation may cause an information and context loss on input data and may adversely effects model performance.

Using keyphrase extraction function to all columns separately:

```
import string
import pke
def keyphrase(text):
    stoplist=list(string.punctuation)
    extractor = pke.unsupervised.YAKE()
    extractor.load_document(input=text,
                           language='en',
                           normalization=None,
                           stoplist=stoplist)
    extractor.candidate_selection(n=5)
    window = 2
    use_stems = False
    extractor.candidate_weighting(window=window,
                                  use_stems=use_stems)
    threshold = 0.8
    keyphrases = extractor.get_n_best(n=3, threshold=threshold)
    return keyphrases[0][0]
```

	Nitel_Soru	Nitel_Cevap	Nitel_Soru	labels
0	Bu göreve gelirken görevinizi yürüten çevre...	Kadınların önünde örnek yok teyzemizden halamı...	Kadınlara rol modeli olabilecek yeterli sayıda...	0
1	Sektörde sektördeki kadın çalışan oranı belirt...	Nihayetinde yönetsel branşlarda fazla kadın ol...	Kadınların okul yöneticisi olarak yeterli oran...	0
2	Sektörde sektördeki kadın çalışan oranı belirt...	Bizde zaten kadın çalışan oranı farklı değil b...	Kadınların okul yöneticisi olarak yeterli oran...	0
3	Kadın yönetici olarak özellikle cinsiyet etmen...	Bir yandan erkek egemen şirketlerde zaman zama...	Kadınlar okul yöneticisi olmak pozitif ayrımcı...	0
4	Kadın yönetici olarak özellikle cinsiyet etmen...	Erkelere verilen imkânlar hiçbir zaman kadına ...	Kadınlar okul yöneticisi olmak pozitif ayrımcı...	0
...
560	Başarılarınızda başarısızlıklarınızda diğer in...	...kimse özgür değildir...	Başarıların başarısızlıkların üzerinde diğer i...	4
561	Başarılarınızda başarısızlıklarınızda diğer in...	...aslında zor felsefi bir soru Akademik çalışma...	Başarıların başarısızlıkların üzerinde diğer i...	4
562	Yaptığınız işin sonuçları üzerinde oynadığınız...	yapmış oldukları işlerin sonuçlarında kendi ko...	Yaptığım işlerin sonuçları üzerinde kontrolümü...	4
563	Başarılarınızda şansın etkisi nedir	...fırsatları iyi görebilmeyi şans olarak görüyo...	Başarıların çoğu şansla ilişkilidir	4
564	Başarılarınızda şansın etkisi nedir	...hayatının seyrini değiştirecek insanlarla tan...	Başarıların çoğu şansla ilişkilidir	4
	Nitel_Soru	Nitel_Cevap	Nitel_Soru	labels
0	yürüten çevrenizden destek aldınız alıyorsunuz	kadınların önünde örnek yok teyzemizden	kadınlara rol modeli olabilecek yeterli	0
1	mevcut sizce kadınların yönetim kademelerinde	nihayetinde yönetsel branşlarda fazla kadın	kadınların okul yöneticisi olarak yeterli	0
2	mevcut sizce kadınların yönetim kademelerinde	bizde zaten kadın çalışan oranı	kadınların okul yöneticisi olarak yeterli	0
3	zorluklar var biraz bahsedebilir misiniz	bir yandan erkek egemen şirketlerde	kadınlar okul yöneticisi olmak pozitif	0
4	zorluklar var biraz bahsedebilir misiniz	erkelere verilen imkânlar hiçbir zaman	kadınlar okul yöneticisi olmak pozitif	0
...
560	başarılarınızda başarısızlıklarınızda diğer in...	kimse özgür değildir	başarıların başarısızlıkların üzerinde diğer i...	4
561	başarılarınızda başarısızlıklarınızda diğer in...	zor felsefi bir soru akademik	başarıların başarısızlıkların üzerinde diğer i...	4
562	yaptığınız işin sonuçları üzerinde oynadığınız	yapmış oldukları işlerin sonuçlarında kendi	yaptığım işlerin sonuçları üzerinde kontrolümün	4
563	başarılarınızda şansın etkisi nedir	ben fırsatları iyi görürüm doğal	başarıların çoğu şansla ilişkilidir	4
564	başarılarınızda şansın etkisi nedir	tanışmamı kesinlikle şansa bağlıyorum şanslıyım	başarıların çoğu şansla ilişkilidir	4

Using keyphrase extraction function after merging columns:

labels		input
0	0	Bu göreve gelirken ve görevinizi yürütürken çe...
1	0	Sektörde . (sektördeki kadın çalışan oranı bel...
2	0	Sektörde . (sektördeki kadın çalışan oranı bel...
3	0	Kadın yönetici olarak özellikle cinsiyet etmen...
4	0	Kadın yönetici olarak özellikle cinsiyet etmen...

labels		input
0	0	aldınız ya da alıyorsunuz
1	0	sektördeki kadın çalışan oranı belirtilecek
2	0	sektördeki kadın çalışan oranı belirtilecek
3	0	kadın yönetici olarak özellikle cinsiyet
4	0	yönetici olarak özellikle cinsiyet etmenini

3.2.2 LIST OF OPERATIONS

As mentioned earlier results which are compared in this study are results of list of operations. Operations consist of various combinations of the functions which are mentioned in functions section. Also to compare with operations which include functions, some of the operations do not include functions, just include cleaned data as input.

Operations in the list of operations are same for each approach in this study. These operations create different input data therefore create different results. Operations in the list are shown below:

- 1-Using cleaned input data without applying functions.
- 2-Applying keyphrase function to the input data.
- 3-Applying both lemmatizer and keyphrase function to the input data
- 4- (only available for single column approach)
- 5- Applying both keyphrase function and lemmatizer function to the merged data (only available for single column approach)

4. RESULTS

Under each approach , operations applied to the input data and different results are obtained for each operation. Results of this study could examine under two main groups which are Results of Five Label approach and Results of Two Label Approach.

Each group has two subgroup which corresponding Single-Column Approach and Multi-Column Approach.

4.1 RESULTS OF FIVE LABEL APPROACH

4.1.1 RESULTS OF SINGLE-COLUMN APPROACH

- 1) Result of the operation which using cleaned input data without applying functions:
- 2) Result of the operation which applying keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	1.433800	1.433098	0.380000
2	1.323700	1.451958	0.500000
3	1.222500	1.376829	0.400000
4	1.137600	1.523781	0.440000
5	0.953800	1.707037	0.420000
6	0.605300	2.076153	0.520000
7	0.336200	2.734418	0.460000
8	0.164800	3.203514	0.460000
9	0.079400	3.203979	0.440000
10	0.046400	3.260996	0.460000

- 3) Result of the operation which applying both lemmatizer and keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	1.444400	1.566565	0.440000
2	1.341500	1.415733	0.420000
3	1.233400	1.336953	0.420000
4	1.088600	1.476428	0.460000
5	0.864400	1.915105	0.440000
6	0.640400	2.325263	0.460000
7	0.403600	2.534942	0.440000
8	0.216100	2.971892	0.500000
9	0.119800	3.068683	0.480000
10	0.067200	3.137208	0.480000

- 4) Result of the operation which applying keyphrase function to the merged data:

Epoch	Training Loss	Validation Loss	Accuracy
1	1.462700	1.579167	0.360000
2	1.355800	1.492323	0.460000
3	1.099300	1.695114	0.360000
4	0.849700	1.975149	0.340000
5	0.694300	2.743031	0.360000
6	0.459700	3.186686	0.400000
7	0.426700	3.364025	0.380000
8	0.338500	3.436075	0.360000
9	0.348300	3.586994	0.380000
10	0.296400	3.706068	0.360000

- 5) Result of the operation which applying both keyphrase function and lemmatizer function to the merged data:

Epoch	Training Loss	Validation Loss	Accuracy
1	1.508100	1.407750	0.360000
2	1.424800	1.395819	0.420000
3	1.230700	1.639558	0.420000
4	0.952500	2.028683	0.380000
5	0.703900	2.436486	0.360000
6	0.574200	2.707069	0.380000
7	0.515400	3.136182	0.360000
8	0.458400	2.813404	0.360000
9	0.378900	3.066464	0.380000
10	0.358800	3.147902	0.360000

4.1.2 RESULTS OF MULTI-COLUMN APPROACH

1) Result of the operation which using cleaned input data without applying functions:

2) Result of the operation which applying keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	1.417600	1.623332	0.360000
2	1.427000	1.517139	0.400000
3	1.263200	1.410673	0.400000
4	1.167900	1.542788	0.440000
5	0.943400	1.937540	0.400000
6	0.649000	2.183258	0.460000
7	0.384600	2.641079	0.440000
8	0.194000	2.989906	0.380000
9	0.065600	3.234350	0.440000
10	0.029800	3.296645	0.380000

3) Result of the operation which applying both lemmatizer and keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	1.421600	1.503545	0.340000
2	1.291900	1.339492	0.500000
3	1.193600	1.357365	0.540000
4	0.997500	1.780798	0.360000
5	0.800300	2.003381	0.420000
6	0.555000	2.513677	0.460000
7	0.302800	3.109448	0.500000
8	0.122900	3.178956	0.500000
9	0.068600	3.174811	0.440000
10	0.041100	3.239058	0.460000

4.2 RESULTS OF A TWO LABEL APPROACH

4.2.1 RESULTS OF SINGLE-COLUMN APPROACH

- 1) Result of the operation which using cleaned input data without applying functions:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.651500	0.607885	0.720000
2	0.623100	0.602733	0.720000
3	0.593400	0.597371	0.720000
4	0.589100	0.599680	0.720000
5	0.590300	0.597584	0.720000
6	0.574400	0.599947	0.720000
7	0.576200	0.608942	0.720000
8	0.583300	0.592940	0.720000
9	0.573300	0.595675	0.720000
10	0.567300	0.594258	0.720000

- 2) Result of the operation which applying keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.682500	0.766444	0.640000
2	0.578100	0.685942	0.640000
3	0.574500	0.740247	0.640000
4	0.628300	0.657424	0.640000
5	0.585400	0.701794	0.640000
6	0.577100	0.672027	0.640000
7	0.565600	0.682302	0.640000
8	0.559400	0.690138	0.640000
9	0.572400	0.683527	0.640000
10	0.561100	0.689177	0.640000

- 3) Result of the operation which applying both lemmatizer and keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.681400	0.725590	0.640000
2	0.586400	0.727197	0.640000
3	0.592800	0.796192	0.640000
4	0.679400	0.658467	0.640000
5	0.579800	0.712626	0.640000
6	0.572400	0.675855	0.640000
7	0.567300	0.682976	0.640000
8	0.558700	0.695826	0.640000
9	0.573300	0.689755	0.640000
10	0.566400	0.691318	0.640000

- 4) Result of the operation which applying keyphrase function to the merged data:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.710700	0.887535	0.640000
2	0.581700	0.963111	0.720000
3	0.526600	1.337133	0.720000
4	0.341500	1.392276	0.720000
5	0.237700	1.757645	0.720000
6	0.233400	1.850313	0.680000
7	0.186600	1.880994	0.680000
8	0.152000	2.189739	0.680000
9	0.169500	1.998355	0.700000
10	0.127000	2.102489	0.700000

- 5) Result of the operation which applying both keyphrase function and lemmatizer function to the merged data:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.698900	0.755227	0.640000
2	0.576800	0.705821	0.640000
3	0.585600	0.733515	0.640000
4	0.634000	0.662611	0.640000
5	0.565200	0.701593	0.640000
6	0.575600	0.679442	0.640000
7	0.568300	0.687078	0.640000
8	0.568900	0.691997	0.640000
9	0.576200	0.686355	0.640000

4.2.2 RESULTS OF A MULTI-COLUMN APPROACH

- 1) Result of the operation which using cleaned input data without applying functions:
- 2) Result of the operation which applying keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.683400	0.810206	0.640000
2	0.579400	0.726246	0.640000
3	0.589000	0.785379	0.640000
4	0.669000	0.665983	0.640000

- 3) Result of the operation which applying both lemmatizer and keyphrase function to the input data:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.677600	0.587310	0.660000
2	0.697100	0.802564	0.780000
3	0.657500	0.805826	0.820000
4	0.464000	0.878725	0.760000
5	0.252400	1.047218	0.780000
6	0.112100	1.144619	0.800000
7	0.063200	1.531078	0.780000
8	0.015800	1.560602	0.780000
9	0.000300	1.628859	0.780000
10	0.009500	1.607256	0.780000