

```
colab.research.google.com/github/ashalogic/Persian-Sentiment-Analyzer/blob/master/Tutorial.ipynb#scrollTo=nS1VtmSF6wmn
Tutorial.ipynb
File Edit View Insert Runtime Tools Help Cannot save changes
+ Code + Text Copy to Drive
[11] WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/python/training/tracking/tracking.py:111: Layer.update
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/python/training/tracking/tracking.py:111: Layer.update
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
INFO:tensorflow:Assets written to: persian-sentiment-fasttext.model/assets

- Step D) A simple form to test our tiny shiny model 😊
there is two form but it's just for showcase there is no diff between them.

[12] user_text: "خلیلی گوشیه خوبیه. تشخیص چهره داره. داخل جعبه کاور گوشیه و محافظ صفحه داره. من دیروز به دستم رسیده عالیه مرسی از دجی کالا"
96 % | 3 %

[17] user_text: "من خیلی بد حال هستم"
1 % | 98 %
```

تشخیص احساسات در نوشتار فارسی (Persian sentiment analysis)

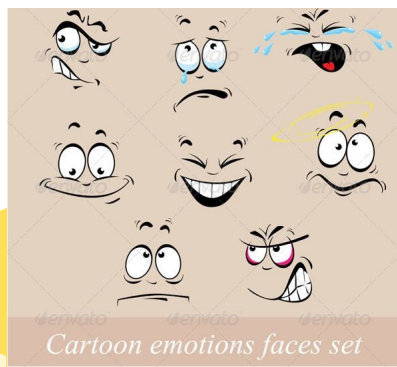
پروژه جالب تشخیص احساسات در متن فارسی (کاری از Parsiya

(Keshavarz

با درود و سپاس

در ادامه آموزش های هوش مصنوعی با گوگل گولب (Google CLoba Notebook) ، در اینجا با معرفی کاری از Parsiya Keshavarz (متن باز در سایت GitHub) ، با شما همراه هستم.

توضیحات مربوط به پروژه را می توانید در ویدیوی زیر مشاهده بفرمایید:



گردآورنده:
سهیل سبزواری

مقدمه

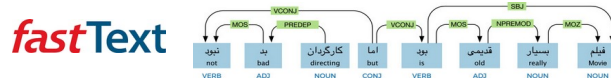
اشتراک دیدگاه ها در وب سایت های ارتباط جمعی، راهی مناسب برای انتقال پیام ها به کاربران نهایی در یک موضوع خاص است .

آنالیز احساسی شاخه ای از پردازش زبان طبیعی (NLP) است، که در زمینه های مختلفی مانند موارد زیر کاربرد دارد :

- (1) بازاریابی
- (2) خدمات مشتری
- (3) و غیره

مدل به کار رفته :

- Hazm برای تحلیل اولیه متن
- fastText (برای نمایش کلمه)
-
- CNN با لایه هایی از LSTM
-
-



نتایج تجربی اثربخشی مدل پیشنهادی را با دقت **۸۵٪** نشان می دهد .

دیتابیس

- مجموعه داده‌های فارسی مختلف را از تویی‌تر و digikala به عنوان مشهورترین سایت دیجیتال جمع‌آوری شده و به ۳ دسته بندی مثبت ، منفی و یا خنثی تقسیم بندی شده (Tutorial_Dataset.csv)
- تصویر زیر بخشی از فایل **دیتابیس** را محیط کولب نشان می دهد:

Tutorial_Dataset.csv		
991 to 1000 of 3261 entries		
Text	Score	Suggestion
من پارسال خریدم اصلا پوشش دهی خوبی نداره نمیدونم فکر کنم به خاطر لایت بودنش چون تا حالا لایت استفاده نکرده بودم چند بار بیشتر استفاده نکردم والا منافذ باز پوست من رو که اصلا کوچک نکرد که بعد از زدن کرم رنگی کمتر به چشم بیان فقط ماندگاری ارایش رو برد بالا کاترینس ازین خیلی بهتره و چروکها و منافذ رو کم میکنه	60	3
سایزش خیلی کوچیکه ولی واقعا عالیه تمام معایب صورت رو میپوشونه منظورم از معایب سوراخ هایی که روی پوست های چربه هست منگه راضی ام ولی ای کاش وزنش هم نوشته میشد	52	3
این کالا رو سفارش دادم ولی وقتی رسید دیدم انقدر براش پرداخت کردم خیلی تو ذوقم خورد چون فکر میکردم ابعادش خیلی بزرگتر باشه کاش دیجی درمورد ابعادشم مینوشت کارایی خاصی ازین ندیدم از خریدش خیلی راضی نیستم ولی اگه کسی به صورت حرفه ای ارایش میکنه شاید بدردش بخوره من از این پنکیک سالهاست که استفاده میکنم و خیلی راضی ام روی پوست خوب میشینه و اصلا جوش نزدنم خیلی خوبه که دیجی کالا اینو آورد خیلی خوشحال شدم و خریدمش	88	1
من از صبح تا شب اینو رو پوستم دارم خیلی سبکه لطیف و مخملیه پوشانندگی بالایی نداره ولی حس خوبی بهت میده پوشش دهیش عالیه ماندگارشم روی پوست چرب خوبه طوری که بدون فیکساتور از صبح تا شب تقریبا ثابت میمونه فقط اگه مدت طولانی قراره بمونه روی پوستتون و پوستتون چربه نباید دست بزنید بهش چون پاک میشه	48	2
واقعا عالی بوی خوبی داره تمیز کنندگی چشم گیری داره چربی پوست رو کم میکنه تو اولین استفاده به وضوح تفاوتش مشخصه	84	1
من این محصول رو برای خرید غروسی ام به صورت خیلی اتفاقی خریدم هیچ بویی نداره اما قدرت پاک کنندگی اش مخصوصا برای ریمل عالیه سخت ترین و سنگین ترین مدادها و ریمل ها و سایه هارو بدون دردسر و راحت پاک میکنه جالب اینه که اگه اتفاقی موقع پاک کردن چشمتان یکم داخل چشم بشه اصلا چشم رو نمی سوزاند و اذیت نمیشود من برای پاک کردن رزلب هم ازین استفاده کردم واسه اونم عالی جواب داد تنها نقطه ضعفش اینه که قبل از استفاده باید تکان اش بدهید و بعدش یکم برنیز روی پنبه و به چشمتان بکشید	100	1
در کل محصول خیلی خوبه فقط من با وجود این که پوست چرب دارم بعد از استفاده هیچ چشمم جتما باید سوراخ ها و منافذ باز منم رو کم نمیشه	64	2
	88	1
	80	1
	96	1

دانلود Databse در فایل Google Colab NoteBooks

Step B) Preparing opinion dataset

in this step we going to collect a dataset that crawled by @minasmz it's not good and I only used 450 pos and 450 neg reviews from it.anyway here we will download the dataset and split it to train and test (I created Train and Test then I filled it with data)

```
#@title Download and prepare Dataset
!wget https://raw.githubusercontent.com/ashalogic/Persian-Sentiment-Analyzer/master/Tutorial_Dataset.csv
```

Tutorial_Dataset.csv

991 to 1000 of 3261 entries

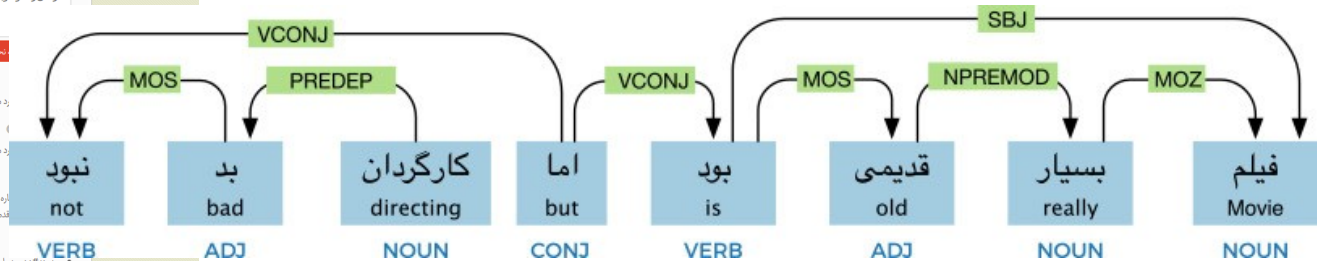
Text	Score	Suggestion
من پارسال خریدم اصل پوشتن دهی خوب نداره نمیدونم فکر کنم به خاطر لایت بودنش چون تا حالا لایت استفاده نکرده بودم چند بار بیشتر استفاده نکردم والا منافذ بار پوشتن من رو که اصلا کوچک نکرد که بعد از ردن کرم رنگی کمتر به چشم بیان فقط ماندگاری ارایشم رو برد بالا کانس آبرن خیلی بهتره و چروکها و منافذ رو کم میکنه	60	3
سایزش خیلی کوچیکه ولی واقعا عالیه تمام معایب صورت رو میوشونه منظورم از معایب سوراخ های که روی پوست های چربه هست منکه راضی ام ولی ای کاش وزنش هم نوشته میشد	52	3
این کالارو سفارش دادم ولی وقتی رسید دیدم انقدر براش پرداخت کردم خیلی تو ذوقم خورد چون فکر میکردم ابعادش خیلی بزرگتر باشه کاش بجای درمورد ابعادش میپوشتن کارایی خاصی ازش ندیدم از خریدش خیلی راضی نیستم ولی اگه کسی به صورت چروقه ای ارایش میکنه شاید بدردش بخوره من از این پنکیک سالهاست که استفاده میکنم و خیلی راضی ام روی پوست خوب میشینه و اصلا جوش نرم خیلی خوبه که بجای کالای ایلو آورد خیلی خوشحال شدم و خریدمش	88	1
من از صبح تا شب اینو رو پوستم دارم خیلی سبکه لطیف و مخمله پوشاندننگی بالایی نداره ولی حس خوبی بهت مده	48	2
پوشتن دهیش عالیه ماندگاریش روی پوست چرب خوبه طوری که بدون فیکساتور از صبح تا شب تقریبا ثابت میمونه فقط اگه مدت طولانی قراره بمونه روی پوستتون و پوستتون چربه باید دست برید بهترش چون پاک میشه	84	1
واقعا عالی بوی خوبی داره تمیز کنندگی چشم گیری داره چربی پوست رو کم میکنه تو اولین استفاده به وضوح تفاوتش مشخصه	100	1
من این محصول رو برای خرید غروب ام به صورت خیلی اتفاقی خریدم هیچ بوی نداره اما قدرت پاک کنندگی اش مخصوصا برای ریمال عالیه سبک ترین و سبکترین موداده و ریمال ها و سایه هارو بدون دردسر و راحت پاک میکنه جالب اینه که اگه اتفاقی موقع پاک کردن چشمتان بکم داخل چشم بشه اصلا چشم رو نم سوزاند و اذیت نمیشود من برای پاک کردن رژب هم ازش استفاده کردم واسه اونم عالی جواب داد تنها نقطه ضعیفش اینه که قبل از استفاده باید تکان اش بدهید و بعدش بکم بریزید روی پنه و به چشمتان بکشید	64	2
در کل محصول خیلی خوبه فقط من با وجود این که پوست چرب دارم بعد از استفاده روی چشمتان حتما باید سریعاً روغن بریم وگرنه میسوره	88	1

Show 10 per page

آماده سازی دیتابیس

- همچنین ۷ در این پروژه با استفاده از ماژول **Hazm** در محیط پایتون ([لینک فرادرس در صورت نیاز](#))، جملات فارسی مانند نمونه زیر تحلیل شده اند و به نحوی برای ورود به عنوان داده های آموزشی و تست در شبکه عصبی استفاده گردیده است:

The screenshot shows the Hazm library website. At the top, there's a header with the Hazm logo and navigation links. Below that, a text input field contains the sentence: "در می‌زند و دوباره وارد میشود. همه می‌گویند قدمش پرپرکت است." To the right of the input field, there's a button labeled "تحلیل". Below the input field, the analysis results are displayed, showing the sentence broken down into words and their corresponding parts of speech (POS) tags. The words are: "ننبود", "بد", "کارگردان", "اما", "بود", "قدیمی", "بسیار", "فیلم". The POS tags are: "not", "bad", "directing", "but", "is", "old", "really", "Movie". The tags are color-coded: VERB (not), ADJ (bad), NOUN (directing), CONJ (but), VERB (is), ADJ (old), NOUN (really), NOUN (Movie). The website also includes a sidebar with links to various resources and a footer with contact information.



آموزش فرادرس در مورد HAZM

بخش Hazm در فایل Google Colab NoteBooks

Installing HAZM

Output of Hazm (Input the CSV Database)

- ▼ Step B) Preparing opinion dataset

in this step we going to collect a dataset that crawled by @minasmz it's not good and I only used 450 pos and 450 neg reviews from it.anyway here we will download the dataset and split it to train and test (I created Train and Test then I filled it with data)

```

!pip install hazm

import random
import numpy
import hazm

def CleanPersianText(text):
    _normalizer = hazm.Normalizer()
    text = _normalizer.normalize(text)
    return text

csv_dataset = pandas.read_csv("/content/Tutorial_Dataset.csv")
revlist = list(map(lambda x: [CleanPersianText(x[0]),x[1]],zip(csv_dataset['Text'],csv_dataset['Suggestion'])))
pos=list(filter(lambda x: x[1] == 1,revlist))
nat=list(filter(lambda x: x[1] == 2,revlist))
neg=list(filter(lambda x: x[1] == 3,revlist))
revlist_shuffle = pos[:450] + neg[:450]
random.shuffle(revlist_shuffle)

print("Posetive count {}".format(len(pos)))
print("Negative count {}".format(len(neg)))
print("Natural count {}".format(len(nat)))
print()
print("Total count {}".format(len(revlist)))
print()

: ", "\n", pos[random.randrange(1, len(pos))])
: ", "\n", neg[random.randrange(1, len(neg))])
: ", "\n", nat[random.randrange(1, len(nat))])
: {}".format(len(revlist_shuffle)))

```

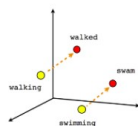
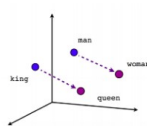
Downlo

Prepare

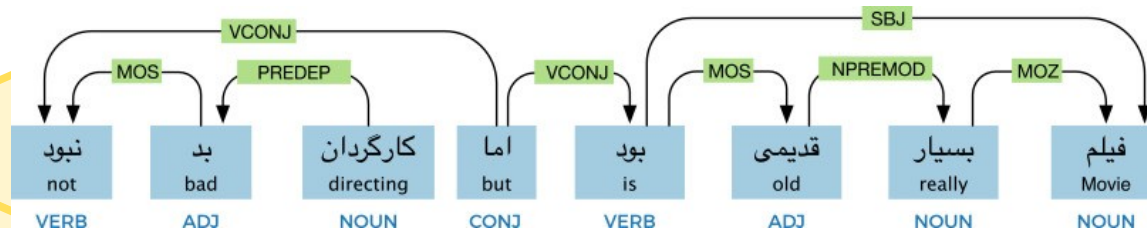
انواع روشهای تعبیه کلمه یا Text Embeddings

- Word2Vec
- Skip-gram Model
- CBOW Model
- GloVe
- fastText

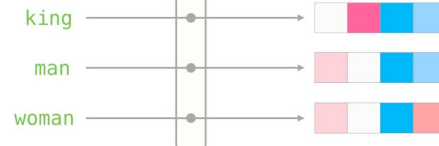
Usage



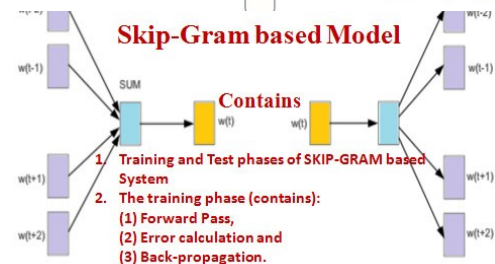
fastText



Word2vec



Skip-gram based Model



دانلود و کار با Fasttext در فایل Google Colab NoteBooks

if you Want to know more then read [this article from Thomas Wolf](#) but now we gonna use Fasttext because it's Pretrained by Facebook and we can use it (there is nothing to worry about this model it's pretty easy to train it by your self or your corpus facebook used Persian Wikipedia and some other staff as dataset for this model so it's just very simpler for us 🤓)

```
#@title Download, extract and load Fasttext word embedding model

!wget https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.fa.300.bin.gz
!gunzip /content/cc.fa.300.bin.gz
!pip install fasttext

import fasttext

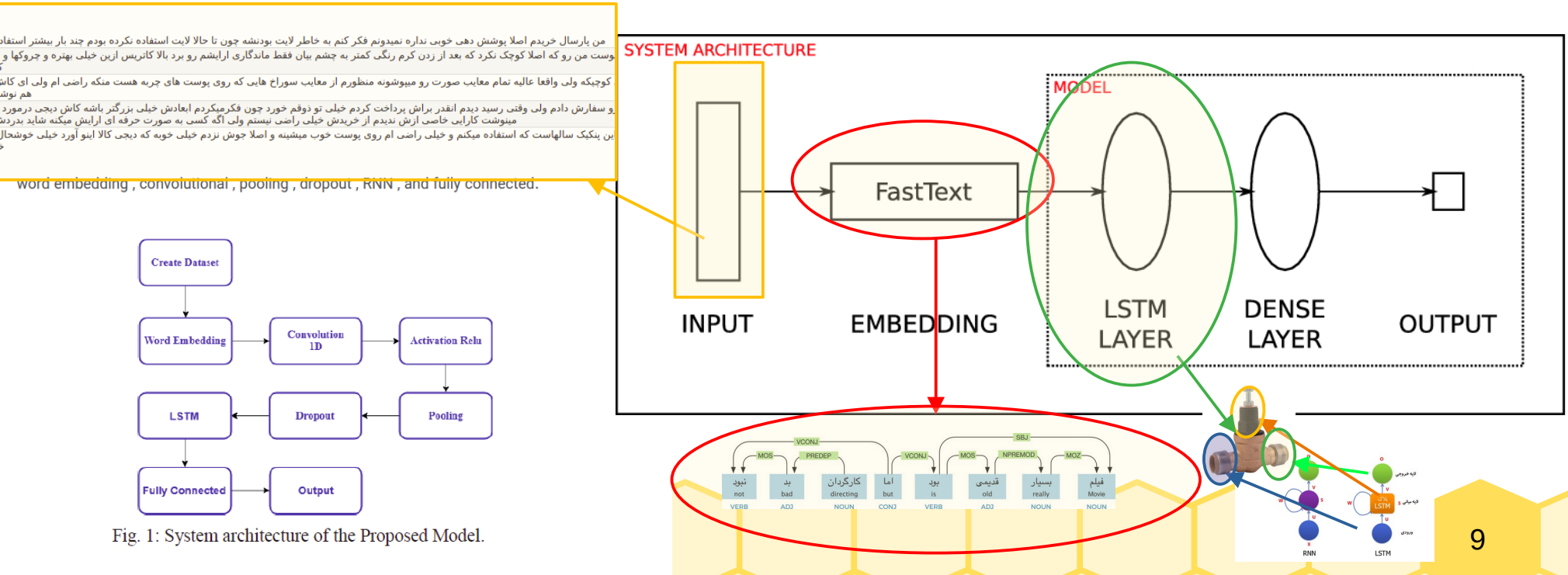
%time
model = fasttext.load_model("/content/cc.fa.300.bin")
```

--2020-10-18 09:09:04-- <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.fa.300.bin.gz>
Resolving dl.fbaipublicfiles.com (dl.fbaipublicfiles.com)... 104.22.75.142, 172.67.9.4, 104.22.74.142, ...
Connecting to dl.fbaipublicfiles.com (dl.fbaipublicfiles.com)|104.22.75.142|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4502524724 (4.2G) [application/octet-stream]
Saving to: 'cc.fa.300.bin.gz'

cc.fa.300.bin.gz 100%[=====>] 4.19G 17.2MB/s in 4m 12s

مدل شبکه عصبی

مراحل انجام کار تقریبا مطابق فلوجارت زیر می باشد:



تنظیمات اولیه شبکه عصبی در فایل Google Colab NoteBooks

Prepare Train & Test Data

تنظیم طول بردار کلمه (vector_size) و
بیشترین تعداد دسته بندی مجاز برای
کلمات (max_no_tokens)

```
▶ #@title Prepare Train & Test Data
vector_size = 300 #@param {type:"integer"}
max_no_tokens = 20 #@param {type:"integer"}
import numpy as np
import keras.backend as K
train_size = int(0.9*(len(revlist_shuffle)))
test_size = int(0.1*(len(revlist_shuffle)))

indexes = set(np.random.choice(len(revlist_shuffle), train_size + test_size, replace=False))

x_train = np.zeros((train_size, max_no_tokens, vector_size), dtype=K.floatx())
y_train = np.zeros((train_size, 2), dtype=np.int32)

x_test = np.zeros((test_size, max_no_tokens, vector_size), dtype=K.floatx())
y_test = np.zeros((test_size, 2), dtype=np.int32)
```

داده های تست و آموزش

```
▶ #@title Prepare Train & Test Data
vector_size = 300 #@param {type:"integer"}
max_no_tokens = 20 #@param {type:"integer"}
import numpy as np
import keras.backend as K
train_size = int(0.9*(len(revlist_shuffle)))
test_size = int(0.1*(len(revlist_shuffle)))
```

90% of Data for Train

10% of Data for test

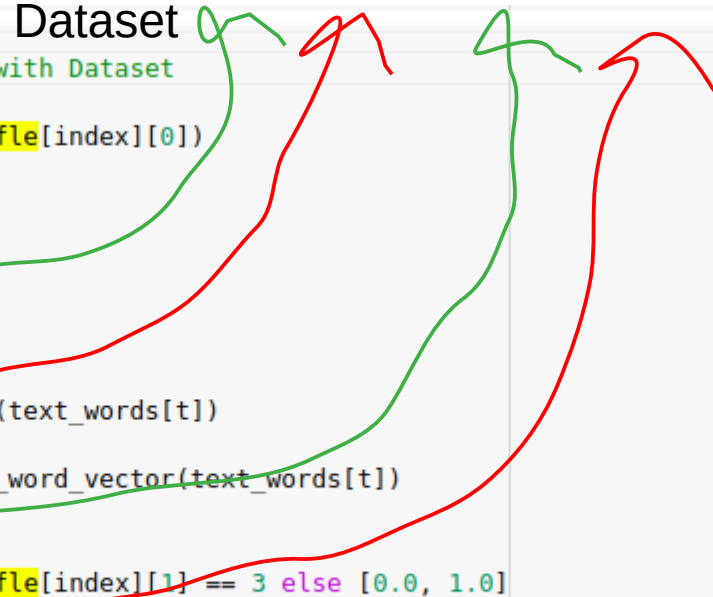
Fill X_Train, X_Test, Y_Train, Y_Test with Dataset

```
#@title Fill X_Train, X_Test, Y_Train, Y_Test with Dataset
for i, index in enumerate(indexes):
    text_words = hazm.word_tokenize(revlist_shuffle[index][0])
    for t in range(0, len(text_words)):
        if t >= max_no_tokens:
            break

        if text_words[t] not in model.words:
            continue
        if i < train_size:
            x_train[i, t, :] = model.get_word_vector(text_words[t])
        else:
            x_test[i - train_size, t, :] = model.get_word_vector(text_words[t])

    if i < train_size:
        y_train[i, :] = [1.0, 0.0] if revlist_shuffle[index][1] == 3 else [0.0, 1.0]
    else:
        y_test[i - train_size, :] = [1.0, 0.0] if revlist_shuffle[index][1] == 3 else [0.0, 1.0]

x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

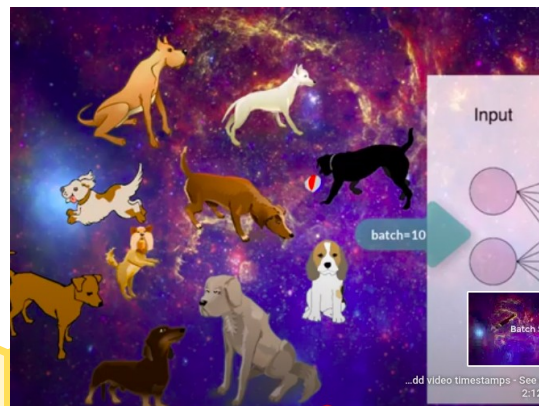
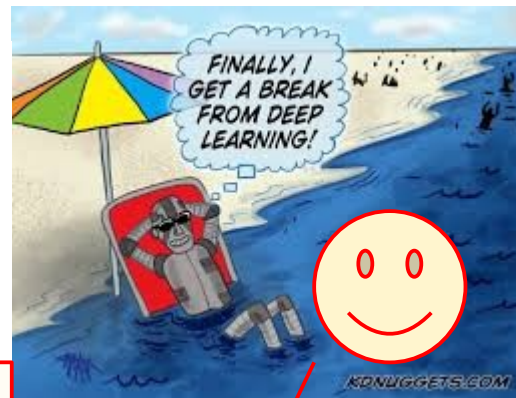
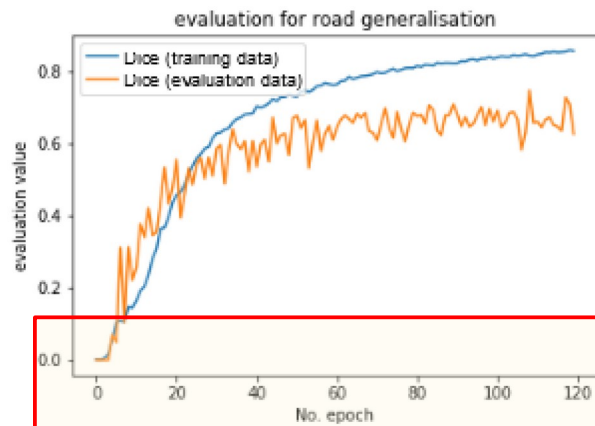


Epochs & batch_size

Step C) Preparing LSTM model

Now we will create our LSTM model then feed it our Train data and b

```
#@title Set batchSize and epochs  
batch_size = 500 #@param {type:"integer"}  
no_epochs = 200 #@param {type:"integer"}  
w2v_model = model  
del model
```



(a)

دوره‌ها (Epochs) نشانگر تعداد

دفعاتی هستند که مجموعه داده

کامل به شبکه عصبی مصنوعی پاس

داده خواهد شد. $batch_size$ تعداد

داده‌های نمونه‌ای است که در یک

دوره به کار می‌رود

لایه های شبکه عصبی

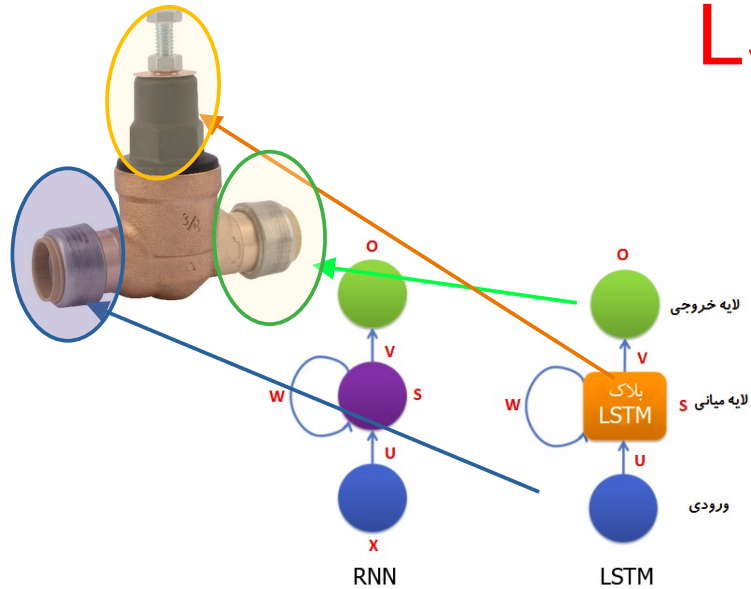
- در این پروژه لایه های متفاوتی به صورت زیر به کار رفته است:
- Conv1D, Dropout, Dense, Flatten, **LSTM**, MaxPooling1D, Bidirectional

```
##title Building LSTM Model
from keras.models import Sequential
from keras.layers import Conv1D, Dropout, Dense, Flatten, LSTM, MaxPooling1D, Bidirectional
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping, TensorBoard
```

Building LSTM Model

یک واحد LSTM

یک واحد LSTM مشترک از موارد زیر میباشد:



(1) یک سلول

(2) یک گیت ورودی

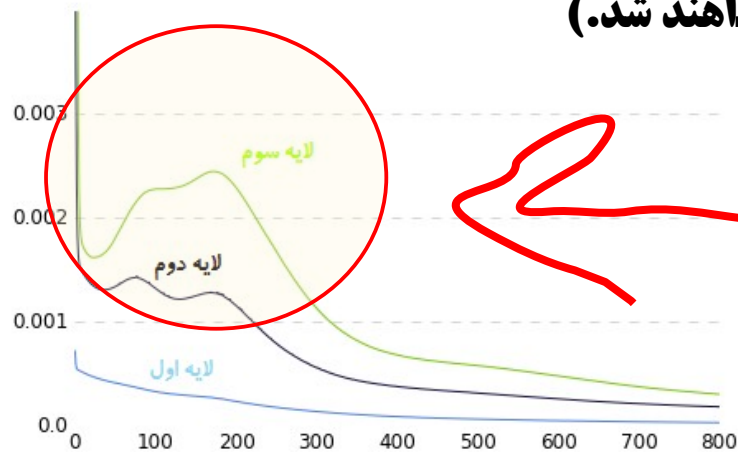
(3) یک گیت فراموشی

(4) یک گیت خروجی

گیت ورودی مشخص می‌کند که کدام اطلاعات باید وارد وضعیت سلول شوند . گیت فراموشی ، مشخص می‌کند که **کدام اطلاعات باید از حالات سلول قبلی حذف شوند** تا فقط اطلاعات مربوطه را حفظ کنند . گیت خروجی مشخص می‌کند که چه مقدار از وضعیت داخلی باید در معرض لایه‌های بالاتر قرار گیرد .

معماری لایه پیچیده LSTM

- در یک شبکه عصبی معمولی تمام ورودی ها و خروجی ها مستقل از یکدیگر هستند. در صورتی که شبکه بتواند روابط بین کلمات را یاد بگیرد مسلماً نمی تواند کلمه بعدی را به درستی پیش بینی کند.
- بزرگترین ویژگی LSTM امکان یادگیری وابستگی بلند مدت است که توسط شبکه های عصبی بازگشتی امکان پذیر نبود.
- امکان وقوع مشکل انفجار گرادیان کماکان در این نوع شبکه ها وجود دارد. (پارمترهای ناصحیح لایه اول منجر به تحت تاثیر قرار گرفتن سایر لایه ها و انفجار گرادیان خواهند شد.)



تنظیمات لایه ها برای شبکه عصبی در فایل Google Colab NoteBooks

+ Code + Text Copy to Drive

no_epochs: 200

```
#@title Building LSTM Model
from keras.models import Sequential
from keras.layers import Conv1D, Dropout, Dense, Flatten, LSTM, MaxPooling1D, Bidirectional
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping, TensorBoard

model = Sequential()

model.add(Conv1D(32, kernel_size=3, activation='elu', padding='same',
                 input_shape=(max_no_tokens, vector_size)))
model.add(Conv1D(32, kernel_size=3, activation='elu', padding='same'))
model.add(Conv1D(32, kernel_size=3, activation='relu', padding='same'))
model.add(MaxPooling1D(pool_size=3))

model.add(Bidirectional(LSTM(512, dropout=0.2, recurrent_dropout=0.3)))

model.add(Dense(512, activation='sigmoid'))
model.add(Dropout(0.2))
model.add(Dense(512, activation='sigmoid'))
model.add(Dropout(0.25))
model.add(Dense(512, activation='sigmoid'))
model.add(Dropout(0.25))

model.add(Dense(2, activation='softmax'))

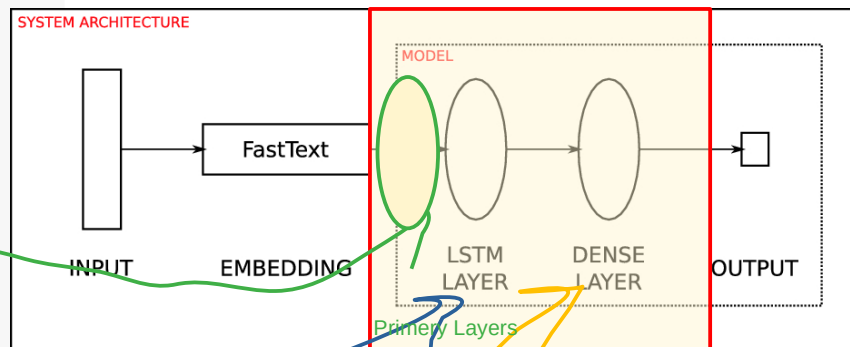
model.compile(loss='categorical_crossentropy', optimizer=Adam(lr=0.0001, decay=1e-6), metrics=['accuracy'])

# tensorboard = TensorBoard(log_dir='logs/', histogram_freq=0, write_graph=True, write_images=True)

model.summary()
```

Building LSTM Model

وارد کردن لایه های مورد استفاده در شبکه عصبی



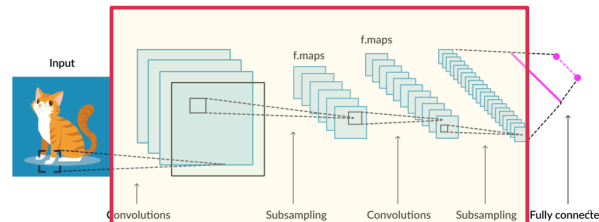
خروجی تنظیمات لایه ها برای شبکه عصبی در فایل Google Colab NoteBooks

WARNING:tensorflow:Layer lstm will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU
WARNING:tensorflow:Layer lstm will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU
WARNING:tensorflow:Layer lstm will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU
Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 20, 32)	28832
conv1d_1 (Conv1D)	(None, 20, 32)	3104
conv1d_2 (Conv1D)	(None, 20, 32)	3104
max_pooling1d (MaxPooling1D)	(None, 6, 32)	0
bidirectional (Bidirectional)	(None, 1024)	2232320
dense (Dense)	(None, 512)	524800
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 512)	262656
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 512)	262656
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 2)	1026
Total params: 3,318,498		
Trainable params: 3,318,498		
Non-trainable params: 0		

لایه های شبکه عصبی به
ترتیب

Output Shape
(None, 20, 32)
(None, 20, 32)
(None, 20, 32)
(None, 6, 32)
(None, 1024)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 2)



ابعاد ماتریس لایه ها

تعداد پارامترهای متغیر در
هر لایه !!

هشدار های قابل
توجه در رابطه با
عدم انجام برخی
تنظیمات و بالا
بودن بار مصرفی
CPU در شبکه

رشد شبکه ساخته شده با داده های تست و آموزش در فایل Google Colab NoteBooks

خروجی اجرای دستورات (قابل توجه
برای تنظیمات بهتر جهت دقت بیشتر
خروجی ها و رفع خطاهای احتمالی)

```
model.fit(x_train, y_train, batch_size=batch_size, shuffle=True, epochs=no_epochs,  
validation_data=(x_test, y_test))
```

```
2/2 [=====] - 0s 59ms/step - loss: 0.0000e+00 - accuracy: 0.4309 - val_loss: 0.0000e+00 - val_accuracy: 0.0000e+00  
Epoch 198/200  
2/2 [=====] - 0s 50ms/step - loss: 0.0000e+00 - accuracy: 0.4519 - val_loss: 0.0000e+00 - val_accuracy: 0.0000e+00  
Epoch 199/200  
2/2 [=====] - 0s 54ms/step - loss: 0.0000e+00 - accuracy: 0.4432 - val_loss: 0.0000e+00 - val_accuracy: 0.0000e+00  
Epoch 200/200  
2/2 [=====] - 0s 48ms/step - loss: 0.0000e+00 - accuracy: 0.4605 - val_loss: 0.0000e+00 - val_accuracy: 0.0000e+00  
<tensorflow.python.keras.callbacks.History at 0x7faf804807f0>
```

تمرین شبکه عصبی با داده های **Train**
(**epochs**, **batch_size**)

```
model.evaluate(x=x_test, y=y_test, batch_size=32, verbose=1)
```

```
3/3 [=====] - 0s 10ms/step - loss: 0.0000e+00 - accuracy: 0.0000e+00  
[0.0, 0.0]
```

تمرین شبکه عصبی با داده های **Test**
(**epochs**, **batch_size**)

```
model.save('persian-sentiment-fasttext.model')
```

ذخیره مدل برای اجرای در زمان ورودی های
جدید مانند درخواست کاربر برای

```
INFO:tensorflow:Assets written to: persian-sentiment-fasttext-model/assets  
INFO:tensorflow:Assets written to: persian-sentiment-fasttext-model/assets  
WARNING:tensorflow:From /usr/local/lib/python3.8/dist-packages/tensorflow/python/ops/resource_variable_ops.py:183: tf.nn.conv2d is deprecated and will be removed in a future version. Use tf.nn.conv2d_v2 instead.  
INFO:tensorflow:Assets written to: persian-sentiment-fasttext-model/assets  
WARNING:tensorflow:From /usr/local/lib/python3.8/dist-packages/tensorflow/python/ops/resource_variable_ops.py:183: tf.nn.conv2d is deprecated and will be removed in a future version. Use tf.nn.conv2d_v2 instead.
```

نتایج و مقایسه

نتایج تجربی اثربخشی مدل پیشنهادی را با دقت **۸۵٪** نشان می‌دهد.

دقت تقریبی مدل های **تشخیص احساسات** در زبان **انگلیسی** با لایه ها و ساختار مشابه:

LSTM FastText

.968

.644

.465

مقایسه با برخی از سایر روشهای بکار رفته در تحلیل احساسات در زبان انگلیسی

جدول زیر مقایسه ای بین دقت مدل‌های مختلف شبکه عصبی برای پردازش کلمات نشان می دهد:

	Travel Scheduling	Ask Ubuntu	Web Applications
NB	.959	.726	.502
SVM FastText average	.958	.812	.771
SVM FastText sum	.968	.800	.658
SVM GloVe average	.946	.805	.591
SVM GloVe sum	.957	.729	.692
SVM Word2Vec average	.979	.742	.698
SVM Word2Vec sum	.946	.742	.680
LSTM FastText	.968	.644	.465
BiLSTM FastText	.979	.646	.549
LSTM GloVe	.945	.665	.546
BiLSTM GloVe	.979	.667	.635
LSTM Word2Vec	.989	.631	.395
BiLSTM Word2Vec	.989	.710	.443
LUIS	.979	.743	.690
Watson	.968	.819	.630
APLai	.931	.782	.628
RASA	.979	.708	.494

Intent Classification for Dialogue Utterances

همچنین امکان استفاده از مدل های دیگر، شبکه عصبی مانند GRU نیز مطابق معماری هایی از قبیل ارایه شده در این مقاله و عکس های زیر، نیز برای این کار، موجود می باشد:

پروژه های مشابه

- A-Determination-of-Effective-Word-Embedding-of-Bengali-Language-using-text-classification

In [0]: bengali_news_after_preprocessing.tail()

Out[0]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
39995	হতভল	অববোধের	করণে	ইউরোপের	ফ্রেডমের	ছোট	আরুত	অন	ফায়র	আরুত	মিডিং	শেষটি	দৈ	বাংলাদেশের	পোশাক	কায়না	পরি
39996	সড়কসড়তে	ছোট	বড়	অসংখ্য	গত	পিচ	চলারিয়ে	অভিদি	ধলাবানু	উভয়ে	সম্মতে	গতি	চলান	দূর	ট্রেট	চলান	করা
39997	মুখিব	বাংলা	দেশ	বেছে	নিয়ন্ত্রিত	কিনগের	বাঙালিয়	মকিন	মজ্জিত	এক	কিনগের	আবেগ	অনুভূতিক	আগম	সাত	বনি	বাং
39998	ফেরদাদি	ইনবিটাস	অজাপক	আসিসজ্ঞানমে	তম	জানবাসিকী	উপসকে	শরিকম	আতীয়	জানুয়ার	মুশ	মিশ্রায়তনে	সংবর্গা	আইলো	সুভেচ্ছা	জায়াতে	আ
39999	টোপা	হতভল	করণে	বতভর	পাইকারি	মোবাম	মহাপাল	বাজার	চাকা	চইগ্রামসর	সারা	দেশে	সবজির	সময়ক	বহু	পাইকারি	বাজ

5 rows x 17 columns

```

1) word_tokenize(text) = list(map(lambda x: word_tokenize(x), word_tokenize(text)))
wordlist = sentence.translate(translate_map).split()
global c, x;
c+=1
if c>=1:
    print(x,end=' ')
    x+=1
return '\n' if (filter(lambda x: x not in stops, wordlist))

```

```

[tk.word_tokenize(text) for text in title]
[sentence to wordlist(text) for text in content]
l.DataFrame(filter(text)
pickle('/content/drive/My Drive/Projects/Bengali Text Classification/Bengali_Text_after_preprocessing.pkl')
# 688 788 888 988 1088 1188 1288 1388 1488 1588 1688 1788 1888 1988 2088 2188 2288 2388 2488 2588 2688 2788 28
188 3288 3388 3488 3588

```

processed pickle text file from drive. Change the path file according to local path file.

```

after preprocessing = pd.read_pickle('/content/drive/My Drive/Projects/Bengali Text Classification/Bengali_Text
rocessing.pkl')

```

after preprocessing.head()

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
১	কর	একজন	গেত	এমন	করা	সকল	কর	করা	করা	করা	করা	করা	করা	করা	করা
২	কর	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা
৩	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা
৪	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা
৫	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা	করা

5 rows x 17 columns

In [0]: bengali_news_after_preprocessing.tail()

Out[0]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
39995	হতভল	অববোধের	করণে	ইউরোপের	ফ্রেডমের	ছোট	আরুত	অন	ফায়র	আরুত	মিডিং	শেষটি	দৈ	বাংলাদেশের	পোশাক	কায়না	পরি
39996	সড়কসড়তে	ছোট	বড়	অসংখ্য	গত	পিচ	চলারিয়ে	অভিদি	ধলাবানু	উভয়ে	সম্মতে	গতি	চলান	দূর	ট্রেট	চলান	করা
39997	মুখিব	বাংলা	দেশ	বেছে	নিয়ন্ত্রিত	কিনগের	বাঙালিয়	মকিন	মজ্জিত	এক	কিনগের	আবেগ	অনুভূতিক	আগম	সাত	বনি	বাং
39998	ফেরদাদি	ইনবিটাস	অজাপক	আসিসজ্ঞানমে	তম	জানবাসিকী	উপসকে	শরিকম	আতীয়	জানুয়ার	মুশ	মিশ্রায়তনে	সংবর্গা	আইলো	সুভেচ্ছা	জায়াতে	আ
39999	টোপা	হতভল	করণে	বতভর	পাইকারি	মোবাম	মহাপাল	বাজার	চাকা	চইগ্রামসর	সারা	দেশে	সবজির	সময়ক	বহু	পাইকারি	বাজ

5 rows x 17 columns

برای مثال

- **GRU** یک مکانیزم **gating** در RNN، شبیه **LSTM**، با پارامترهای کمتر - <آموزش آسان می باشد.

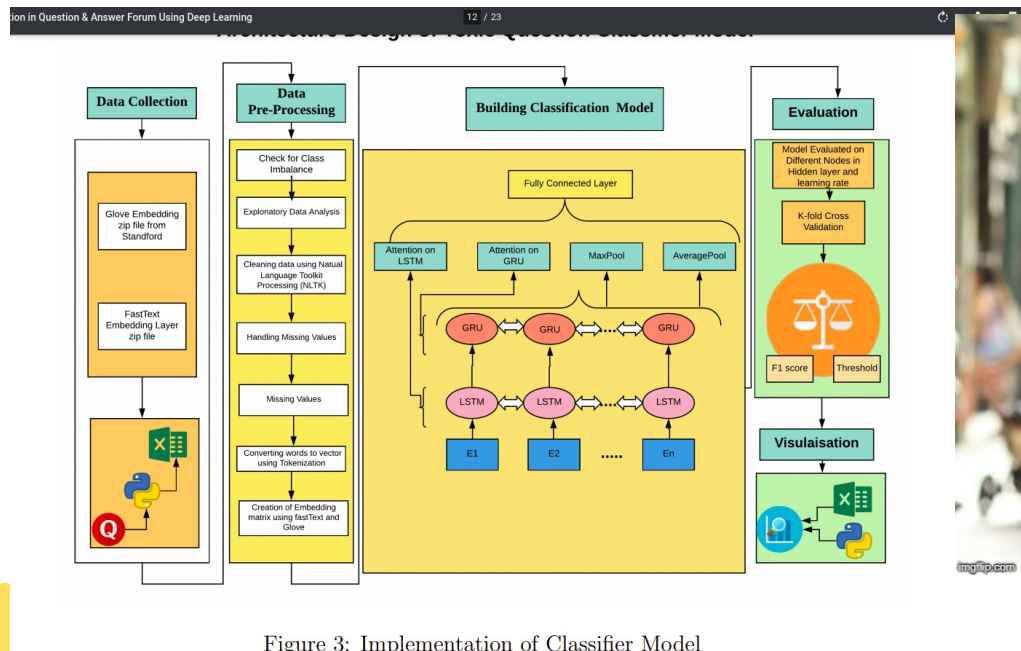
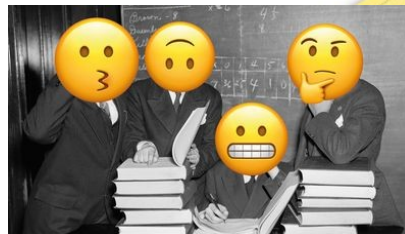


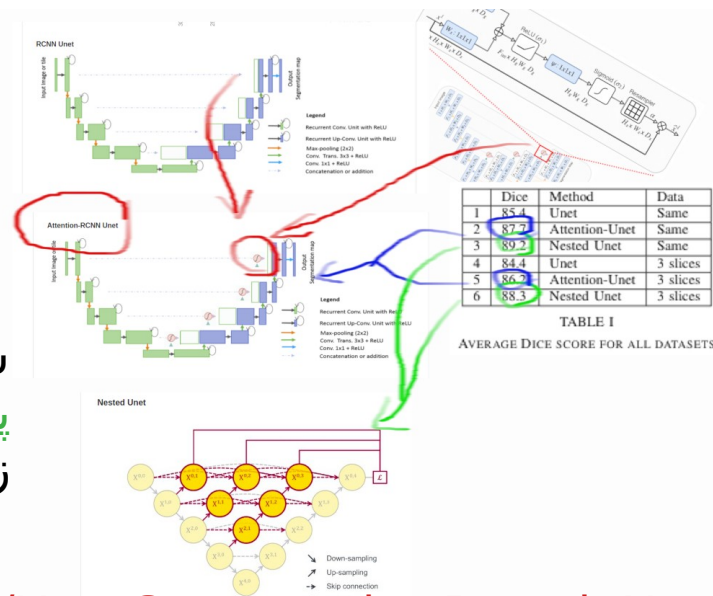
Figure 3: Implementation of Classifier Model

Some Strange suggestion!!

- Based of these two Github project;
 - [Image_Segmentation](#)
 - [Unet-Segmentation-Pytorch-Nest-of-Unets](#)
- It could be tested by some more accuracte candidate methods like :
 - [Attention-RCNN Unet](#)
 - [Nested Unet](#)
 - ['R2AttU_Net'](#)
 - ['R2AttU_Net'](#)



سوال : ایا استفاده از مدل های جدید در زمینه
پردازش تصویر (مانند مدل های این اسلاید) در
زمینه پردازش زبان رقابتی خواهد بود؟



برخی منابع مرتبط، فارسی جالب توجه

- یک برنامه وب شباهت یاب متن با TensorFlow.js بسازید

Persian Sentiment Analysis With LSTM & Fasttext

یک برنامه وب شباهت یاب متن با TensorFlow.js بسازید

شباهت کسینوسی



شباهت کسینوسی با محاسبه زاویه کسینوسی بین دو بردار، یک معیار شباهت است. اگر دو بردار شباهت باشند، زاویه بین آن‌ها کوچک است و مقدار نشانده کسینوسی به 1 نزدیکتر است.



پروژه جالب تشخیص احساسات در متن فارسی (کاری از Parsiya و Keshavarz)

با توجه به منابع در دسترس، برای آموزش مدل‌های LSTM و GRU (Google Colab Notebook) در اینجا با معرفی کردن

توضیحات مربوط به پروژه را می‌توانید در وبسایت زیر مشاهده نمایید:



پروژه جالب تشخیص احساسات در متن فارسی

Step D) A simple form to test our tiny shiny model 🌟

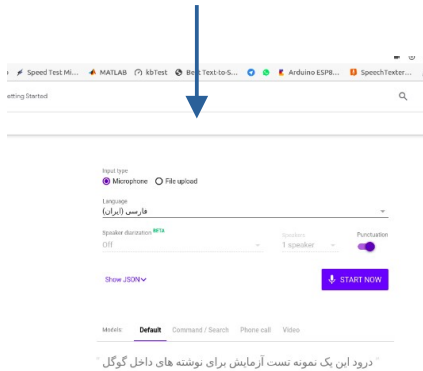
there is two form but it's just for showcase there is no diff between them.

- user_text: جایی گوشه خویه، تشخیص چهره داره. داخل جعبه کاور گوشی و محافظ صفحه داره. من دیروز به دستم رسیده عالیه مرسی از دجی کالا
- user_text: جایی جالبه این موبایل اصلا همه چی تمامه من که پسندیدم این موبایل ریا رو

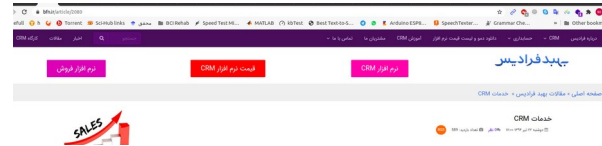
برخی الگوهای قابل توسعه

<https://cloud.google.com/speech-to-text/?hl=en>

- استفاده از تبدیل صوت به متن گوگل (تست و هزینه ها در لینک بالا) با هزینه پایین برای کاربردهایی از قبیل :



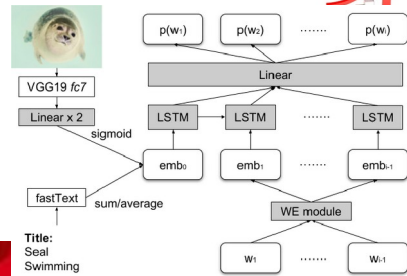
<https://www.bfn.ir/article/2080>



... خدمات CRM

ترکیب پردازش تصویر با پردازش متن

بهبود قوانین بر اساس شادی و غم



تحلیل دادهای برخی آزمایشات تجربی !!



در پایان



با تشکر از توجه و نظرات شما برای ...