```
In [1]: # importing libraries
        import warnings
        warnings.filterwarnings("ignore")
        import pandas as pd
        import numpy as np
        from keras.layers import Input, Embedding, LSTM, Dropout, BatchNormaliz
        ation, Dense, concatenate, Flatten, Conv1D, MaxPool1D, LeakyReLU, ELU,
        SpatialDropout1D, MaxPooling1D, GlobalAveragePooling1D, GlobalMaxPoolin
        q1D
        from keras.preprocessing.text import Tokenizer, one hot
        from keras.preprocessing.sequence import pad sequences
        from keras.models import Model, load model
        from keras import regularizers
        from keras.optimizers import *
        from keras.callbacks import ModelCheckpoint, EarlyStopping, TensorBoard
         , ReduceLROnPlateau
        from sklearn.feature extraction.text import TfidfVectorizer, CountVecto
        rizer
        from sklearn.metrics import roc auc score
        import tensorflow as tf
        from tensorboardcolab import *
        import matplotlib.pyplot as plt
        %matplotlib inline
        import re
        from tadm import tadm
        from sklearn.preprocessing import LabelEncoder
        import seaborn as sns
        import pickle
        Using TensorFlow backend.
In [2]: # https://medium.com/@rushic24/mounting-google-drive-in-google-colab-5e
        cd1d3b735a
        # https://towardsdatascience.com/3-ways-to-load-csv-files-into-colab-7c
        14fcbdcb92#targetText=To%20start%2C%20log%20into%20your,Colab%20has%20i
```

```
t%20installed%20already).
from google.colab import drive
drive.mount('/content/drive')
```

Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth? client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleuser content.com&redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Aoob&scope=email%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdocs.test%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdrive%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdrive.photos.readonly%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fpeopleapi.readonly&response type=code

Enter your authorization code:
.....
Mounted at /content/drive

In [3]: # reading datasets project_data = pd.read_csv("/content/drive/My Drive/Data/preprocessed_d ata.csv") #pd.read_csv("preprocessed_data.csv") project_data.head()

Out[3]:

s	chool_state	teacher_prefix	project_grade_category	teacher_number_of_previously_p		
O ca	a	mrs	grades_prek_2	53		
1 u	ıt	ms	grades_3_5	4		

		L		<u>L</u>		
	school_state	teacher_prefix	project_grade_category	teacher_number_of_previously_p		
2	са	mrs	grades_prek_2	10		
3	ga	mrs	grades_prek_2	2		
4	wa	mrs	grades_3_5	2		
4 ▮				>		
<pre>print("Number of data points in train data", project_data.shape) print('-'*50) print("The attributes of data :", project_data.columns.values)</pre>						
Number of data points in train data (109248, 9)						
The attributes of data : ['school_state' 'teacher_prefix' 'project_grad e_category' 'teacher_number_of_previously_posted_projects' 'project_is_approved' 'clean_categories' 'clean_subcategories' 'essay' 'price']						

In [4]:

```
In [5]: approved_project = project_data['project_is_approved'].values
    project_data.drop(['project_is_approved'], axis=1, inplace=True)
    project_data.head(1)
```

Out[5]:

	school_state teacher_prefix r		project_grade_category	teacher_number_of_previously_p		
O	ca	mrs	grades_prek_2	53		

In [0]: # Data splitting

from sklearn.model_selection import train_test_split

Splitting in train and test

X_train, X_test, y_train, y_test = train_test_split(project_data, appro ved_project, test_size=0.33, stratify=approved_project)

In [7]: #clearing the graph of tensorflow
 tf.keras.backend.clear session()

input_seq_total_text_data = Input(shape=(300,),name="input_seq_total_te
xt data")

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:541: The name tf.placeholder is deprecated. Please use tf.compat.v1.placeholder instead.

In [8]: # https://machinelearningmastery.com/use-word-embedding-layers-deep-lea
rning-keras/

```
# prepare tokenizer
         text tokenizer = Tokenizer()
         text_tokenizer.fit_on_texts(X_train["essay"])
         vocab size = len(text tokenizer.word index) + 1
         vocab size
Out[8]: 48319
In [0]: # integer encode the data
         encoded essay train = text tokenizer.texts to sequences(X train["essay"
         encoded essay test = text tokenizer.texts to sequences(X test["essay"])
In [10]: # Padding data
         padded text train = pad sequences(encoded essay train, maxlen=300, padd
         ing='post', truncating='post')
         padded text test = pad sequences(encoded essay test, maxlen=300, paddin
         g='post', truncating='post')
         print(padded text train.shape)
         print(padded text test.shape)
         (73196, 300)
         (36052, 300)
In [0]: f = open("/content/drive/My Drive/Data/glove vectors", "rb")
         glove words = pickle.load(f)
In [12]: # create a weight matrix for words in training docs
         embedding matrix = np.zeros((vocab size, 300))
         for word, i in text tokenizer.word index.items():
             embedding vector = glove words.get(word)
             if embedding vector is not None:
```

embedding_matrix[i] = embedding_vector
print(embedding_matrix.shape)

(48319, 300)

In [13]: Emb_Txt_Data = Embedding(vocab_size, 300, weights = [embedding_matrix],
 input_length = 300, trainable=False)

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:66: The name tf.get_default_graph is deprecated. Please use tf.compat.v1.get default graph instead.

In [14]: Emb_Text_Data = (Emb_Txt_Data)(input_seq_total_text_data)

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:4432: The name tf.random_uniform is deprecated. Please use tf.random.uniform instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:190: The name tf.get_default_session is deprecated. Please use tf.compat.v1.get_default_session instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:197: The name tf.ConfigProto is deprecated. Please use tf.compat.v1.ConfigProto instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:203: The name tf.Session is deprecated. Ple ase use tf.compat.v1.Session instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:207: The name tf.global_variables is deprecated. Please use tf.compat.v1.global_variables instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:216: The name tf.is_variable_initialized is

deprecated. Please use tf.compat.v1.is_variable_initialized instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:223: The name tf.variables_initializer is deprecated. Please use tf.compat.v1.variables initializer instead.

```
In [15]: Emb Text Data
Out[15]: <tf.Tensor 'embedding 1/embedding lookup/Identity:0' shape=(?, 300, 30</pre>
         0) dtype=float32>
In [16]: | lstm = LSTM(64, recurrent dropout=0.5, kernel regularizer = regularizer
         s.l2(0.001), return sequences = True)(Emb Text Data)
         WARNING: tensorflow: From /usr/local/lib/python3.6/dist-packages/keras/ba
         ckend/tensorflow backend.py:148: The name tf.placeholder with default i
         s deprecated. Please use tf.compat.vl.placeholder with default instead.
         WARNING: tensorflow: From /usr/local/lib/python3.6/dist-packages/keras/ba
         ckend/tensorflow backend.py:3733: calling dropout (from tensorflow.pyth
         on.ops.nn ops) with keep prob is deprecated and will be removed in a fu
         ture version.
         Instructions for updating:
         Please use `rate` instead of `keep prob`. Rate should be set to `rate =
         1 - keep prob`.
In [17]: | lstm
Out[17]: <tf.Tensor 'lstm 1/transpose 1:0' shape=(?, ?, 64) dtype=float32>
In [0]: flatten = Flatten()(lstm)
In [19]: flatten
Out[19]: <tf.Tensor 'flatten 1/Reshape:0' shape=(?, ?) dtype=float32>
In [0]: # processing for school state
```

```
input school state = Input(shape=(50,),name="input school state")
In [21]:
         unique school state = X train["school state"].nunique()
         print(unique school state)
         Emb State Data = Embedding(unique school state, 300, input length = 50)
         (input school state)
         51
In [0]: flatten 1 = Flatten()(Emb State Data)
In [0]: encoded school state train = [one hot(d, unique school state) for d in
         X train['school state']]
         encoded school state test = [one hot(d, unique school state) for d in X
          test['school state']]
         padded school state train = pad sequences(encoded school state train, m
In [24]:
         axlen=50, padding='post')
         padded school state test = pad sequences(encoded_school_state_test, max
         len=50, padding='post')
         print(padded school state train.shape)
         print(padded school state test.shape)
         (73196, 50)
         (36052, 50)
In [25]: # processing for project grade category
         input project grade category = Input(shape=(50,),name="input project gr
         ade category")
         unique project grade = X train["project grade category"].nunique()
         print(unique project grade)
         Emb PGC Data = Embedding(unique project grade, 300, input length = 50)(
```

```
input project grade category)
         flatten 2 = Flatten()(Emb PGC Data)
         4
In [26]: encoded_project_grade_train = [one_hot(d, unique_project grade) for d i
         n X train['project grade category']]
         encoded project grade test = [one hot(d, unique project grade) for d in
          X test['project grade category']]
         padded project grade train = pad sequences(encoded project grade train,
          maxlen=50, padding='post')
         padded project grade test = pad sequences(encoded project grade test, m
         axlen=50, padding='post')
         print(padded project grade train.shape)
         print(padded project grade test.shape)
         (73196, 50)
         (36052, 50)
In [27]: # processing for clean categories
         input clean categories = Input(shape=(50,),name="input clean categorie")
         s")
         unique clean categories = X train["clean categories"].nunique()
         print(unique clean categories)
         Emb clean categories Data = Embedding(unique clean categories, 300, inp
         ut length = 50)(input clean categories)
         flatten 3 = Flatten()(Emb clean categories Data)
         51
In [28]: encoded clean categories train = [one hot(d, unique clean categories) f
         or d in X train['clean categories']]
         encoded clean categories test = [one hot(d, unique clean categories) fo
         r d in X test['clean categories']]
```

```
padded_clean_categories_train = pad sequences(encoded clean categories
         train, maxlen=50, padding='post')
         padded clean categories test = pad sequences(encoded clean categories t
         est, maxlen=50, padding='post')
         print(padded clean categories train.shape)
         print(padded clean categories test.shape)
         (73196, 50)
         (36052, 50)
In [29]: # processing for clean subcategories
         input clean subcategories = Input(shape=(50,),name="input clean subcate
         gories")
         unique clean subcategories = X train["clean subcategories"].nunique()
         print(unique clean subcategories)
         Emb clean subcategories Data = Embedding(unique clean subcategories, 30
         0, input length = 50)(input clean subcategories)
         flatten 4 = Flatten()(Emb clean subcategories Data)
         394
         encoded clean subcategories train = [one hot(d, unique clean subcategor
         ies) for d in X train['clean subcategories']]
         encoded clean subcategories test = [one hot(d, unique clean subcategori
         es) for d in X test['clean subcategories']]
         padded clean subcategories train = pad seguences(encoded clean subcateg
         ories train, maxlen=50, padding='post')
         padded clean subcategories test = pad sequences(encoded clean subcatego
         ries test, maxlen=50, padding='post')
         print(padded clean subcategories train.shape)
         print(padded clean subcategories test.shape)
```

```
(73196, 50)
         (36052, 50)
In [31]: # processing for teacher prefix
         input teacher prefix = Input(shape=(50,),name="input teacher prefix")
         unique teacher prefix = X train["teacher prefix"].nunique()
         print(unique teacher prefix)
         Emb teacher prefix Data = Embedding(unique teacher prefix, 300, input l
         ength = 50)(input teacher prefix)
         flatten 5 = Flatten()(Emb teacher prefix Data)
         5
In [32]: encoded teacher prefix train = [one hot(d, unique teacher prefix) for d
          in X train['teacher prefix']]
         encoded teacher prefix test = [one hot(d, unique teacher prefix) for d
         in X test['teacher prefix']]
         padded teacher prefix train = pad sequences(encoded teacher prefix trai
         n, maxlen=50, padding='post')
         padded teacher prefix test = pad sequences(encoded teacher prefix test,
          maxlen=50, padding='post')
         print(padded teacher prefix train.shape)
         print(padded teacher prefix test.shape)
         (73196, 50)
         (36052, 50)
In [33]: # Processing numerical features
         input numerical data = Input(shape=(2,),name="input numerical data")
         Dense for rem input = Dense(units=32,activation='relu',kernel initializ
         er='he normal',name="Dense for rem input")(input numerical data)
```

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:4479: The name tf.truncated_normal is deprecated. Please use tf.random.truncated normal instead.

```
In [0]: teacher number of previously posted projects train = X train['teacher n
         umber of previously posted projects'].values.reshape(-1, 1)
         price train = X train['price'].values.reshape(-1, 1)
         teacher number of previously posted projects test = X test['teacher num
         ber of previously posted projects'].values.reshape(-1, 1)
         price test = X test['price'].values.reshape(-1, 1)
         concat numerical train = np.concatenate((teacher number of previously p
         osted projects train,price train),axis=1)
         concat numerical test = np.concatenate((teacher number of previously po
         sted projects test,price test),axis=1)
In [0]: concatenate = concatenate(inputs=[flatten, flatten 1, flatten 2, flatte
         n 3, flatten 4, flatten 5, input numerical data], name="concatenate")
In [36]: Dense layer1 after concat = Dense(128,activation="relu", kernel initial
         izer="he normal", kernel regularizer=regularizers.l2(0.001))(concatenat
         e)
         dropout = Dropout(0.5)(Dense layer1 after concat)
         Dense layer2 after concat = Dense(64,activation="relu", kernel initiali
         zer="he normal", kernel regularizer=regularizers.l2(0.001))(dropout)
         dropuout 1 = Dropout(0.5)(Dense layer2 after concat)
         Dense layer3 after concat = Dense(32,activation='relu',kernel initializ
         er='he normal', kernel regularizer=regularizers.l2(0.001))(dropuout_1)
         Dense layer4 after concat = Dense(16,activation='relu',kernel initializ
         er='he normal', kernel regularizer=regularizers.l2(0.001))(Dense layer3
         after concat)
```

```
output layer to classify with softmax = Dense(2,activation='softmax',ke
rnel initializer="he normal", name="output")(Dense layer4 after concat)
model 1 = Model(inputs=[input seq total text data,input school state,in
put project grade category, input clean categories, input clean subcatego
ries,input teacher prefix,input numerical data],outputs=[output layer t
o classify with softmax])
model 1.summary()
Model: "model 1"
Layer (type)
                                Output Shape
                                                     Param #
                                                                  Connec
ted to
input seq total text data (Inpu (None, 300)
                                                      0
embedding 1 (Embedding)
                                (None, 300, 300)
                                                     14495700
                                                                  input
seq total text data[0][0]
input school state (InputLayer) (None, 50)
                                                      0
input project grade category (I (None, 50)
input clean categories (InputLa (None, 50)
input clean subcategories (Inpu (None, 50)
```

<pre>input_teacher_prefix (InputLaye</pre>	(None,	50)	0	
lstm_1 (LSTM) ing_1[0][0]	(None,	300, 64)	93440	embedd
embedding_2 (Embedding) school_state[0][0]	(None,	50, 300)	15300	input_
embedding_3 (Embedding) project_grade_category[0][0	(None,	50, 300)	1200	input_
embedding_4 (Embedding) clean_categories[0][0]	(None,	50, 300)	15300	input_
embedding_5 (Embedding) clean_subcategories[0][0]	(None,	50, 300)	118200	input_
embedding_6 (Embedding) teacher_prefix[0][0]	(None,	50, 300)	1500	input_
flatten_1 (Flatten) [0][0]	(None,	19200)	0	lstm_1
flatten_2 (Flatten) ing_2[0][0]	(None,	15000)	0	embedd
flatten_3 (Flatten)	(None,	15000)	0	embedd

/None 120\

Ω

40000

drapaut 1 (Drapaut)

	aropout_1 (Dropout) 	(None,	120)	ט	uense_	
	dense_2 (Dense) t_1[0][0]	(None,	64)	8256	dropou	
	dropout_2 (Dropout) 2[0][0]	(None,	64)	0	dense_	
	dense_3 (Dense) t_2[0][0]	(None,	32)	2080	dropou	
	dense_4 (Dense) 3[0][0]	(None,	16)	528	dense_	
	output (Dense) 4[0][0]	(None,	2)	34	dense_	
	Total params: 26,809,522 Trainable params: 12,313,822 Non-trainable params: 14,495,700					
In [37]:	<pre># https://github.com/taomanwai/tensorboardcolab/blob/master/README.md tbc=TensorBoardColab()</pre>					
	Wait for 8 seconds TensorBoard link: https://6291a79d.ngrok.io					

```
In [0]: # https://machinelearningmastery.com/check-point-deep-learning-models-k
         eras/
         # https://machinelearningmastery.com/how-to-stop-training-deep-neural-n
         etworks-at-the-right-time-using-early-stopping/
         # https://medium.com/singlestone/keras-callbacks-monitor-and-improve-yo
         ur-deep-learning-205a8a27e91c
         # https://www.tensorflow.org/tensorboard/get started
         # https://keras.rstudio.com/reference/callback tensorboard.html
         # https://colab.research.google.com/drive/lafN2SALDooZIHbBGmWZMT6cZ8ccV
         ElWk#scrollTo=4pxUfiLhbS4Y&forceEdit=true&sandboxMode=true
         #tensorboard model 1 = TensorBoard(log dir='./log', histogram freq=1, w
         rite graph=True, write grads=True, batch size=512, write images=True)
         #callbacks 1 = [tensorboard model 1]
In [0]: model 1 train data = [padded text train,padded school state train,padde
         d project grade train, padded clean categories train, padded clean subcat
         egories train, padded teacher prefix train, concat numerical train]
         model 1 test data = [padded text test,padded school state test,padded p
         roject grade test, padded clean categories test, padded clean subcategori
         es test, padded teacher prefix test, concat numerical test]
In [0]: from keras.utils import np utils
         Y train = np utils.to categorical(y train, 2)
         Y test = np utils.to categorical(y test, 2)
In [0]: # https://stackoverflow.com/questions/41032551/how-to-compute-receiving
         -operating-characteristic-roc-and-auc-in-keras
         def auroc(y true, y pred):
             return tf.py func(roc auc score, (y true, y pred), tf.double)
In [42]: model 1.compile(optimizer='adam', loss='categorical crossentropy', metr
         ics=[auroc])
```

WARNING:tensortlow:From /usr/local/lib/python3.6/dist-packages/keras/op timizers.py:793: The name tf.train.Optimizer is deprecated. Please use tf.compat.vl.train.Optimizer instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:3576: The name tf.log is deprecated. Please use tf.math.log instead.

WARNING:tensorflow:From <ipython-input-41-2303c0155baf>:3: py_func (from tensorflow.python.ops.script_ops) is deprecated and will be removed in a future version.

Instructions for updating:

- tf.py_func is deprecated in TF V2. Instead, there are two options available in V2.
- tf.py_function takes a python function which manipulates tf eager tensors instead of numpy arrays. It's easy to convert a tf eager tensor to

an ndarray (just call tensor.numpy()) but having access to eager tensors

means `tf.py_function`s can use accelerators such as GPUs as well a
s

being differentiable using a gradient tape.

- tf.numpy_function maintains the semantics of the deprecated tf.py func
- (it is not differentiable, and manipulates numpy arrays). It drops the

stateful argument making all functions stateful.

In [43]: history = model_1.fit(model_1_train_data, Y_train, batch_size=512, epoc
hs=20, verbose=1, validation_data=(model_1_test_data, Y_test), callback
s=[TensorBoardColabCallback(tbc)])

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow_core/python/ops/math_grad.py:1424: where (from tensorflow.python.op s.array_ops) is deprecated and will be removed in a future version. Instructions for updating:

Use tf.where in 2.0, which has the same broadcast rule as np.where WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow backend.py:1033: The name tf.assign add is deprecated.

Please use tf.compat.vl.assign add instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:1020: The name tf.assign is deprecated. Ple ase use tf.compat.v1.assign instead.

Train on 73196 samples, validate on 36052 samples WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorbo ardcolab/core.py:49: The name tf.summary.FileWriter is deprecated. Plea se use tf.compat.v1.summary.FileWriter instead.

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/callbacks.py:1122: The name tf.summary.merge_all is deprecated. Please us e tf.compat.v1.summary.merge_all instead.

Epoch 2/20 5497 - auroc: 0.6927 - val loss: 0.5446 - val auroc: 0.7346 Epoch 3/20 5031 - auroc: 0.7088 - val loss: 0.5081 - val auroc: 0.7440 Epoch 4/20 4750 - auroc: 0.7186 - val loss: 0.4785 - val auroc: 0.7344 Epoch 5/20 4534 - auroc: 0.7271 - val loss: 0.4611 - val auroc: 0.7533 Epoch 6/20 4377 - auroc: 0.7346 - val loss: 0.4467 - val auroc: 0.7550 Epoch 7/20

```
4262 - auroc: 0.7366 - val loss: 0.4237 - val auroc: 0.7522
Epoch 8/20
4170 - auroc: 0.7409 - val loss: 0.4248 - val auroc: 0.7555
Epoch 9/20
4114 - auroc: 0.7412 - val loss: 0.4039 - val auroc: 0.7612
Epoch 10/20
4054 - auroc: 0.7444 - val loss: 0.3970 - val auroc: 0.7632
Epoch 11/20
4018 - auroc: 0.7459 - val loss: 0.3979 - val auroc: 0.7577
Epoch 12/20
3989 - auroc: 0.7474 - val loss: 0.3933 - val auroc: 0.7650
Epoch 13/20
3971 - auroc: 0.7471 - val loss: 0.3901 - val auroc: 0.7621
Epoch 14/20
3939 - auroc: 0.7492 - val loss: 0.3873 - val auroc: 0.7652
Epoch 15/20
3925 - auroc: 0.7478 - val loss: 0.3908 - val auroc: 0.7662
Epoch 16/20
3912 - auroc: 0.7508 - val loss: 0.3871 - val auroc: 0.7665
Epoch 17/20
3910 - auroc: 0.7513 - val loss: 0.3839 - val auroc: 0.7653
Epoch 18/20
3877 - auroc: 0.7538 - val loss: 0.3810 - val auroc: 0.7681
Epoch 19/20
3888 - auroc: 0.7512 - val loss: 0.3833 - val auroc: 0.7653
Epoch 20/20
```

7.5

5.0

10.0

12.5

Train ROCAUC Test ROCAUC

17.5

15.0

0.700

0.675

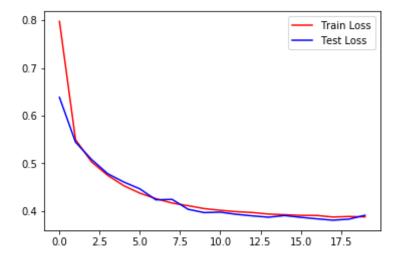
0.650

0.625

0.600

0.0

2.5



Model ROCAUC Value is 0.7672.