

DonorsChoose.org receives hundreds of thousands of project proposals each year for classroom projects in need of funding. Right now, a large number of volunteers is needed to manually screen each submission before it's approved to be posted on the DonorsChoose.org website.

Next year, DonorsChoose.org expects to receive close to 500,000 project proposals. As a result, there are three main problems they need to solve:

- How to scale current manual processes and resources to screen 500,000 projects so that they can be posted as quickly and as efficiently as possible
- How to increase the consistency of project vetting across different volunteers to improve the experience for teachers
- How to focus volunteer time on the applications that need the most assistance

The goal of the competition is to predict whether or not a DonorsChoose.org project proposal submitted by a teacher will be approved, using the text of project descriptions as well as additional metadata about the project, teacher, and school. DonorsChoose.org can then use this information to identify projects most likely to need further review before approval.

# DonorsChoose

## About the DonorsChoose Data Set

The `train.csv` data set provided by DonorsChoose contains the following features:

Feature		Description
<code>project_id</code>		A unique identifier for the proposed project. <b>Example:</b> p036502
<code>project_title</code>	<ul style="list-style-type: none"><li>•</li><li>•</li></ul>	Title of the project. <b>Examples:</b> <code>Art Will Make You Happy!</code> <code>First Grade Fun</code>
<code>project_grade_category</code>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>	Grade level of students for which the project is targeted. One of the following enumerated values: <code>Grades PreK-2</code> <code>Grades 3-5</code> <code>Grades 6-8</code> <code>Grades 9-12</code>
<code>project_subject_categories</code>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li></ul>	One or more (comma-separated) subject categories for the project from the following enumerated list of values: <code>Applied Learning</code> <code>Care &amp; Hunger</code> <code>Health &amp; Sports</code> <code>History &amp; Civics</code> <code>Literacy &amp; Language</code> <code>Math &amp; Science</code> <code>Music &amp; The Arts</code> <code>Special Needs</code> <code>Warmth</code>  <b>Examples:</b> <ul style="list-style-type: none"><li>• <code>Music &amp; The Arts</code></li><li>• <code>Literacy &amp; Language, Math &amp; Science</code></li></ul>
<code>school_state</code>		State where school is located ( <a href="#">Two-letter U.S. postal code</a> ). <b>Example:</b> WY
<code>project_subject_subcategories</code>	<ul style="list-style-type: none"><li>•</li><li>•</li></ul>	One or more (comma-separated) subject subcategories for the project. <b>Examples:</b> <code>Literacy</code> <code>Literature &amp; Writing, Social Sciences</code>
<code>project_resource_summary</code>	<ul style="list-style-type: none"><li>•</li></ul>	An explanation of the resources needed for the project. <b>Example:</b> <code>My students need hands on literacy materials to manage sensory needs!</code>
<code>project_essay_1</code>		First application essay*
<code>project_essay_2</code>		Second application essay*
<code>project_essay_3</code>		Third application essay*

Feature	Description
project_essay_4	Fourth application essay
project_submitted_datetime	Datetime when project application was submitted. <b>Example:</b> 2016-04-28 12:43:56.245
teacher_id	A unique identifier for the teacher of the proposed project. <b>Example:</b> bdf8baa8fedef6bfeec7ae4ff1c15c56
teacher_prefix	Teacher's title. One of the following enumerated values: <ul style="list-style-type: none"> <li>nan</li> <li>Dr.</li> <li>Mr.</li> <li>Mrs.</li> <li>Ms.</li> <li>Teacher.</li> </ul>
teacher_number_of_previously_posted_projects	Number of project applications previously submitted by the same teacher. <b>Example:</b> 2

\* See the section **Notes on the Essay Data** for more details about these features.

Additionally, the `resources.csv` data set provides more data about the resources required for each project. Each line in this file represents a resource required by a project:

Feature	Description
id	A <code>project_id</code> value from the <code>train.csv</code> file. <b>Example:</b> p036502
description	Description of the resource. <b>Example:</b> Tenor Saxophone Reeds, Box of 25
quantity	Quantity of the resource required. <b>Example:</b> 3
price	Price of the resource required. <b>Example:</b> 9.95

**Note:** Many projects require multiple resources. The `id` value corresponds to a `project_id` in `train.csv`, so you use it as a key to retrieve all resources needed for a project:

The data set contains the following label (the value you will attempt to predict):

Label	Description
project_is_approved	A binary flag indicating whether DonorsChoose approved the project. A value of 0 indicates the project was not approved, and a value of 1 indicates the project was approved.

## Notes on the Essay Data

Prior to May 17, 2016, the prompts for the essays were as follows:

- \_\_project\_essay\_1\_\_: "Introduce us to your classroom"
- \_\_project\_essay\_2\_\_: "Tell us more about your students"
- \_\_project\_essay\_3\_\_: "Describe how your students will use the materials you're requesting"
- \_\_project\_essay\_3\_\_: "Close by sharing why your project will make a difference"

Starting on May 17, 2016, the number of essays was reduced from 4 to 2, and the prompts for the first 2 essays were changed to the following:

- \_\_project\_essay\_1\_\_: "Describe your students: What makes your students special? Specific details about their background, your neighborhood, and your school are all helpful."
- \_\_project\_essay\_2\_\_: "About your project: How will these materials make a difference in your students' learning and improve their school lives?"

For all projects with `project_submitted_datetime` of 2016-05-17 and later, the values of `project_essay_3` and `project_essay_4` will be NaN.

In [0]:

```
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")

import sqlite3
import pandas as pd
import numpy as np
import nltk
import math
import string
```

```

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import confusion_matrix
from sklearn import metrics
from sklearn.metrics import roc_curve, auc
from nltk.stem.porter import PorterStemmer

import re
# Tutorial about Python regular expressions: https://pymotw.com/2/re/
import string
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.stem.wordnet import WordNetLemmatizer

from gensim.models import Word2Vec
from gensim.models import KeyedVectors
import pickle

from tqdm import tqdm
import os

from plotly import plotly
import plotly.offline as offline
import plotly.graph_objs as go
offline.init_notebook_mode()
from collections import Counter

from sklearn.tree import DecisionTreeClassifier

import dill #To store session variables
#https://stackoverflow.com/questions/34342155/how-to-pickle-or-store-jupyter-ipython-notebook-session-for-later

```

## 1.1 Reading Data

In [0]:

```

from google.colab import drive
drive.mount('/content/drive', force_remount=True)

```

Go to this URL in a browser: [https://accounts.google.com/o/oauth2/auth?client\\_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&redirect\\_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Ab%3Ascope=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](https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Ab%3Ascope=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 [0]:

```
ls "drive/My Drive/Colab Notebooks"
```

```

'06 Implement SGD.ipynb'          DT_Graph.PNG
3_DonorsChoose_KNN_final.ipynb   glove.6B.50d.txt
4_DonorsChoose_NB_final.ipynb    glove_vectors_300d
5_DonorsChoose_LR_final.ipynb    glove_vectors_50
7_DonorsChoose_SVM_final.ipynb   knn.sess
7_DonorsChoose_SVM.ipynb         resources.csv
8_DonorsChoose_DT_final.ipynb    'SQL Assignment.ipynb'
8_DonorsChoose_DT.ipynb          train_data.csv
'Copy of Welcome To Colaboratory' Untitled0.ipynb
Db-IMDB.db

```

In [0]:

```
project_data = pd.read_csv('drive/My Drive/Colab Notebooks/train_data.csv')
resource_data = pd.read_csv('drive/My Drive/Colab Notebooks/resources.csv')
```

In [0]:

```
project_data_1=project_data[project_data['project_is_approved']==1]
project_data_0=project_data[project_data['project_is_approved']==0]

print(project_data_1.shape)
print(project_data_0.shape)

#Creating a dataset of 0.2k points containg points from both the classes
project_data = project_data_1[0:33458].append(project_data_0[0:16542])
print(project_data['project_is_approved'].value_counts())
print(project_data.shape)
```

```
(92706, 17)
(16542, 17)
1    33458
0    16542
Name: project_is_approved, dtype: int64
(50000, 17)
```

In [0]:

```
print("Number of data points in train data", project_data.shape)
print('-'*50)
print("The attributes of data :", project_data.columns.values)
```

```
Number of data points in train data (50000, 17)
-----
The attributes of data : ['Unnamed: 0' 'id' 'teacher_id' 'teacher_prefix' 'school_state'
 'project_submitted_datetime' 'project_grade_category'
 'project_subject_categories' 'project_subject_subcategories'
 'project_title' 'project_essay_1' 'project_essay_2' 'project_essay_3'
 'project_essay_4' 'project_resource_summary'
 'teacher_number_of_previously_posted_projects' 'project_is_approved']
```

In [0]:

```
# how to replace elements in list python: https://stackoverflow.com/a/2582163/4084039
cols = ['Date' if x=='project_submitted_datetime' else x for x in list(project_data.columns)]

#sort dataframe based on time pandas python: https://stackoverflow.com/a/49702492/4084039
project_data['Date'] = pd.to_datetime(project_data['project_submitted_datetime'])
project_data.drop('project_submitted_datetime', axis=1, inplace=True)
project_data.sort_values(by=['Date'], inplace=True)

# how to reorder columns pandas python: https://stackoverflow.com/a/13148611/4084039
project_data = project_data[cols]

project_data.head(2)
```

Out[0]:

	Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_s
473	100660	p234804	cbc0e38f522143b86d372f8b43d4cff3	Mrs.	GA	2016-04-27 00:53:00	Grades PreK-2	
29891	146723	p099708	c0a28c79fe8ad5810da49de47b3fb491	Mrs.	CA	2016-04-27 01:10:09	Grades 3-5	Math

In [0]:

```
print("Number of data points in train data", resource_data.shape)
print(resource_data.columns.values)
```

```
print(resource_data.columns.values,
resource_data.head(2))
```

Number of data points in train data (1541272, 4)  
['id' 'description' 'quantity' 'price']

Out[0]:

	id	description	quantity	price
0	p233245	LC652 - Lakeshore Double-Space Mobile Drying Rack	1	149.00
1	p069063	Bouncy Bands for Desks (Blue support pipes)	3	14.95

## 1.2 preprocessing of project\_subject\_categories

In [0]:

```
categories = list(project_data['project_subject_categories'].values)
# remove special characters from list of strings python:
https://stackoverflow.com/a/47301924/4084039

# https://www.geeksforgeeks.org/removing-stop-words-nltk-python/
# https://stackoverflow.com/questions/23669024/how-to-strip-a-specific-word-from-a-string
# https://stackoverflow.com/questions/8270092/remove-all-whitespace-in-a-string-in-python
cat_list = []
for i in categories:
    temp = ""
    # consider we have text like this "Math & Science, Warmth, Care & Hunger"
    for j in i.split(','): # it will split it in three parts ["Math & Science", "Warmth", "Care & Hunger"]
        if 'The' in j.split(): # this will split each of the category based on space "Math & Science" => "Math", "&", "Science"
            j = j.replace('The', '') # if we have the words "The" we are going to replace it with '' (i.e removing 'The')
            j = j.replace(' ', '') # we are replacing all the ' ' (space) with '' (empty) ex: "Math & Science" => "Math&Science"
            temp += j.strip() + " " # " abc ".strip() will return "abc", remove the trailing spaces
            temp = temp.replace('&', '_') # we are replacing the & value into
    cat_list.append(temp.strip())

project_data['clean_categories'] = cat_list
project_data.drop(['project_subject_categories'], axis=1, inplace=True)

from collections import Counter
my_counter = Counter()
for word in project_data['clean_categories'].values:
    my_counter.update(word.split())

cat_dict = dict(my_counter)
sorted_cat_dict = dict(sorted(cat_dict.items(), key=lambda kv: kv[1]))
```

## 1.3 preprocessing of project\_subject\_subcategories

In [0]:

```
sub_categories = list(project_data['project_subject_subcategories'].values)
# remove special characters from list of strings python:
https://stackoverflow.com/a/47301924/4084039

# https://www.geeksforgeeks.org/removing-stop-words-nltk-python/
# https://stackoverflow.com/questions/23669024/how-to-strip-a-specific-word-from-a-string
# https://stackoverflow.com/questions/8270092/remove-all-whitespace-in-a-string-in-python

sub_cat_list = []
for i in sub_categories:
    temp = ""
    # consider we have text like this "Math & Science, Warmth, Care & Hunger"
    for j in i.split(','): # it will split it in three parts ["Math & Science", "Warmth", "Care & Hunger"]
```

```

    if 'The' in j.split(): # this will split each of the category based on space "Math & Science"
        e=> "Math","&", "Science"
        j=j.replace('The','') # if we have the words "The" we are going to replace it with '' (i.e removing 'The')
        j = j.replace(' ','') # we are replacing all the ' ' (space) with '' (empty) ex: "Math & Science" => "Math&Science"
        temp +=j.strip()+" #" abc ".strip() will return "abc", remove the trailing spaces
        temp = temp.replace('&','_')
        sub_cat_list.append(temp.strip())

project_data['clean_subcategories'] = sub_cat_list
project_data.drop(['project_subject_subcategories'], axis=1, inplace=True)

# count of all the words in corpus python: https://stackoverflow.com/a/22898595/4084039
my_counter = Counter()
for word in project_data['clean_subcategories'].values:
    my_counter.update(word.split())

sub_cat_dict = dict(my_counter)
sorted_sub_cat_dict = dict(sorted(sub_cat_dict.items(), key=lambda kv: kv[1]))

```

## 1.3 Text preprocessing

In [0]:

```

# merge two column text dataframe:
project_data["essay"] = project_data["project_essay_1"].map(str) + \
    project_data["project_essay_2"].map(str) + \
    project_data["project_essay_3"].map(str) + \
    project_data["project_essay_4"].map(str)

```

In [0]:

```
project_data.head(2)
```

Out[0]:

Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_grade_category
473	100660 p234804	cbc0e38f522143b86d372f8b43d4cff3	Mrs.	GA	2016-04-27 00:53:00	Grades PreK-2	Flex Seating Flex Learning
29891	146723 p099708	c0a28c79fe8ad5810da49de47b3fb491	Mrs.	CA	2016-04-27 01:10:09	Grades 3-5	Brea Box to Learning Engager

In [0]:

```

# printing some random reviews
print(project_data['essay'].values[0])
print("="*50)
print(project_data['essay'].values[150])
print("="*50)
print(project_data['essay'].values[1000])

```

I recently read an article about giving students a choice about how they learn. We already set goals; why not let them choose where to sit, and give them options of what to sit on? I teach at a low-income (Title 1) school. Every year, I have a class with a range of abilities, yet they are all the same age. They learn differently, and they have different interests. Some have ADHD, and some are fast learners. Yet they are eager and active learners that want and need to be able to move around the room, yet have a place that they can be comfortable to complete their work. We need a classroom rug that we can use as a class for reading time, and students can use during other learning times. I have also requested four Kore Kids wobble chairs and four Back Jack padded portable chairs so that students can still move during whole group lessons without disrupting the class. Having th

These areas will provide these little ones with a way to wiggle while working. Benjamin Franklin once said, "Tell me and I forget, teach me and I may remember, involve me and I learn." I want these children to be involved in their learning by having a choice on where to sit and how to learn, all by giving them options for comfortable flexible seating.

=====

A unit that has captivated my students and one that has forced them to seek out further resources on their own, is the Holocaust unit. This unit not only brought their critical thinking skills to life, but it brought out their passion, love, dislikes, and fears about wars and prejudices to light. My 8th graders students live in a high-poverty school district and live in a large, urban area. They are reluctant readers unless introduced to life-changing books. This book made my students work hard in improving their reading and writing skills. The Holocaust unit brought compassion and history to life. The students wanted to read ahead and learn about tolerance and discrimination. These materials will be used in-class. We were read, discuss, and think critically about the world event that still affects us. The Holocaust is part of our history and its victims and survivors deserve our knowledge and recognition of the hardships they endured. We will be researching the victims and survivors of the Holocaust, read non-fictional text, watch documentaries, and overall broaden our education on this historic event. This project will greatly benefit my students. It will not only help them academically and help prepare them for high school, but it will make them well-rounded individuals who better understand the power of tolerance and war. Please know that you have made a positive impact on my students and we sincerely thank you in advance.

=====

Why learn coding in the 5th grade? I teach science through STEM. Instead of using only spaghetti and marshmallows for engineering, I want the students to use coding. It is time to use interactive approaches to solving problems and testing ideas using real-life skills students may use in the future. My school is located in Jupiter, Florida, and we are an intermediate center, servicing only 3rd-5th grades. I teach 3 classes of science to 5th grade students. My students are a mix of gifted and advanced 10 and 11 year olds, of at which 20% have some type of learning challenge, such as ADHD or autism. They all have insatiable thirsts for science. Most come to me with limited knowledge of science, but a tremendous understanding of technology. Most have a computer in their home and are familiar with tablets and smartphones. At least 1/3 of my students know Scratch and JavaScript programming. My goal is to pair my students incredible knowledge of technology with science concepts to deepen their understandings of that concept. I also want to expose all of my students with coding since research has shown that more computer coders will be needed for future jobs than ever before. What I envision is the students working in groups using the specific coding device, Raspberry Pi, to create codes to manipulate the sensors. These will be attached to laptops at each table. In the beginning, I will use the device to teach basic coding to solve a problem. The students will be required to learn how to set up the motherboard during this process. Then I will move on to using it with my science content. One activity I found intriguing is the weather station sensors. The students work together to find a way to code for each of these sensors to turn on and off and collect, store, and manipulate the data. This will become a part of my weather unit. By pairing this type of technology with science, I feel my lesson then is reflecting how science works in the real world. Technology and science go hand in hand and I want my students to experience that one influences the other. I want them to experience that scientists use technology as a tool to further deepen their understanding of concepts. I also want both my boys and girls to learn and understand coding as a viable future career.

In [0]:

```
# https://stackoverflow.com/a/47091490/4084039
import re

def decontracted(phrase):
    # specific
    phrase = re.sub(r"won't", "will not", phrase)
    phrase = re.sub(r"can't", "can not", phrase)

    # general
    phrase = re.sub(r"n't", " not", phrase)
    phrase = re.sub(r"\'re", " are", phrase)
    phrase = re.sub(r"\'s", " is", phrase)
    phrase = re.sub(r"\'d", " would", phrase)
    phrase = re.sub(r"\'ll", " will", phrase)
    phrase = re.sub(r"\'t", " not", phrase)
    phrase = re.sub(r"\'ve", " have", phrase)
    phrase = re.sub(r"\'m", " am", phrase)
    return phrase
```

In [0]:

```
sent = decontracted(project_data['essay'].values[2000])
print(sent)
print("="*50)
```

My school is in a low socio-economic area with a high ELL population. The students in my classroom

=====

```
# \r \n \t remove from string python: http://texthandler.com/info/remove-line-breaks-python/
sent = sent.replace('\r', ' ')
sent = sent.replace('\n', ' ')
sent = sent.replace('\t', ' ')
print(sent)
```

```
#remove spacial character: https://stackoverflow.com/a/5843547/4084039
sent = re.sub('[^A-Za-z0-9]+', ' ', sent)
print(sent)
```

```
# https://gist.github.com/steeleier/554280
# we are removing the words from the stop words list: 'no', 'nor', 'not'
stopwords= ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've",
\
            "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his',
'himself', \
            'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them',
'their', \
            'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll",
'these', 'those', \
            'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having',
'do', 'does', \
            'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', '
while', 'of', \
            'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during',
'before', 'after', \
            'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under'
, 'again', 'further', \
```



```

'ach', 'few', 'more', \
'most', 'other', 'some', 'such', 'only', 'own', 'same', 'so', 'than', 'too', 'very', \
's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll'
, 'm', 'o', 're', \
've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "do
esn't", 'hadn', \
"hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn',
"mightn't", 'mustn', \
"mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn',
"wasn't", 'weren', "weren't", \
'won', "won't", 'wouldn', "wouldn't"]

```

In [0]:

```

# Combining all the above stundents
from tqdm import tqdm
preprocessed_essays = []
# tqdm is for printing the status bar
for sentence in tqdm(project_data['essay'].values):
    sent = decontracted(sentence)
    sent = sent.replace('\\r', ' ')
    sent = sent.replace('\\\"', ' ')
    sent = sent.replace('\\n', ' ')
    sent = re.sub('[^A-Za-z0-9]+', ' ', sent)
    # https://gist.github.com/sebleier/554280
    sent = ' '.join(e for e in sent.split() if e.lower() not in stopwords)
    preprocessed_essays.append(sent.lower().strip())

```

100%|██████████| 50000/50000 [00:24<00:00, 2006.46it/s]

In [0]:

```

#adding a new column for the processed essay text
project_data['clean_essay']=preprocessed_essays
print(project_data.columns)

# after preprocesing
preprocessed_essays[2000]

```

```

Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects', 'project_is_approved',
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay'],
      dtype='object')

```

Out[0]:

'school low socio economic area high ell population students classroom not lot academic practice outside school day love coming school everyday eager learn work hard excited master new concepts school site strive make every minute school day order ensure students able learn feel successful know time precious asking mini white boards reusable write wipe pockets order help monitor students thinking learning often times work done worksheets feedback students not meaningful take awhile give student individual feedback white boards write wipe pockets give students way show written responses gathered carpet together allow give immediate feedback students modify responses right lead meaningful learning processing nannan'

## 1.4.1 Preprocessing of `project\_title`

In [0]:

```
project_data.head(2)
```

Out[0]:

Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_title
------------	----	------------	----------------	--------------	------	------------------------	---------------

Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_title
473	100660	p234804	cbc0e38f522143b86d372f8b43d4cff3	Mrs.	GA	2016-04-27 00:53:00	Grades PreK-2

29891	146723	p099708	c0a28c79fe8ad5810da49de47b3fb491	Mrs.	CA	2016-04-27 01:10:09	Grades 3-5
-------	--------	---------	----------------------------------	------	----	---------------------	------------



In [0]:

```
#Printing a few random review summaries

for i in range(1,3000,1000):
    sent = project_data['project_title'].values[i]
    print(sent, '--- Row No:', i)
    print("="*50)
```

```
Breakout Box to Ignite Engagement! --- Row No: 1
=====
Cozy Classroom Carpet for Learning --- Row No: 1001
=====
Community Circle Carpet: A Place to Call Home! --- Row No: 2001
=====
```

In [0]:

```
# The above random records show that there are no URLs or HTML tags, but we will remove incase if there are any

from tqdm import tqdm #for status bar
from bs4 import BeautifulSoup #for html tags

preprocessed_title=[]

for title in tqdm(project_data['project_title'].values):
    # To remove urls - https://stackoverflow.com/a/40823105/4084039
    title = re.sub(r"http\S+", "", title)

    # To remove all HTML tags
    #https://stackoverflow.com/questions/16206380/python-beautifulsoup-how-to-remove-all-tags-from-an-element
    title = BeautifulSoup(title, 'lxml').get_text()

    # To split contractions - refer decontracted function defined above
    title = decontracted(title)

    # To remove alphanumerics (words with numbers in them) -
    #https://stackoverflow.com/a/18082370/4084039
    title = re.sub("\S*\d\S*", "", title).strip()

    # To remove special characters - https://stackoverflow.com/a/5843547/4084039
    title = re.sub('[^A-Za-z]+', '', title)

    # To remove stop words from the summaries and convert to lowercase
    title = ' '.join(e.lower() for e in title.split() if e.lower() not in stopwords)
    preprocessed_title.append(title.strip())

#adding a new column for cleaned titles
project_data['clean_title']=preprocessed_title
print(project_data.columns)
```

100%|██████████| 50000/50000 [00:11<00:00, 4364.22it/s]

```
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      ...])
```

```
'teacher_number_of_previously_posted_projects', 'project_is_approved',  
'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',  
'clean_title'],  
dtype='object')
```

## 1.4.2 Preprocessing of `teacher\_prefix`

In [0]:

```
#replacing Nan values with 'Unknown'  
project_data['teacher_prefix']=project_data['teacher_prefix'].replace(np.nan, 'Unknown')
```

## 1.4.3 Combining resource\_data with project\_data

In [0]:

```
price_data = resource_data.groupby('id').agg({'price':'sum', 'quantity':'sum'}).reset_index()  
project_data = pd.merge(project_data, price_data, on='id', how='left')
```

## 1.4.4 Adding word counts for Title and Essay

In [0]:

```
#https://stackoverflow.com/questions/54397096/how-to-do-word-count-on-pandas-dataframe  
  
project_data['title_wc'] = project_data['clean_title'].str.count(' ')+1  
  
project_data['essay_wc'] = project_data['clean_essay'].str.count(' ')+1  
  
project_data.columns
```

Out[0]:

```
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',  
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',  
      'project_essay_2', 'project_essay_3', 'project_essay_4',  
      'project_resource_summary',  
      'teacher_number_of_previously_posted_projects', 'project_is_approved',  
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',  
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc'],  
      dtype='object')
```

## 1.4.5 Adding sentiment scores for each essay

In [0]:

```
#http://t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html  
  
import nltk  
from nltk.sentiment.vader import SentimentIntensityAnalyzer  
  
nltk.download('vader_lexicon')  
  
project_data['senti_score'] = 0  
project_data['senti_score'] = project_data['senti_score'].astype(float)  
  
anlyzr = SentimentIntensityAnalyzer()  
  
for index in project_data.index:  
    project_data.at[index, 'senti_score'] = anlyzr.polarity_scores(project_data.at[index, 'clean_essay'])['compound']  
  
print(project_data.columns)
```

```
/usr/local/lib/python3.6/dist-packages/nltk/twitter/__init__.py:20: UserWarning:
```

```
The twython library has not been installed. Some functionality from the twitter package will not be available.
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects', 'project_is_approved',
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc',
      'senti_score'],
      dtype='object')
```

## 1.5 Preparing data for models

```
In [0]:
```

```
project_data.columns
```

```
Out[0]:
```

```
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects', 'project_is_approved',
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc',
      'senti_score'],
      dtype='object')
```

we are going to consider

- school\_state : categorical data
- clean\_categories : categorical data
- clean\_subcategories : categorical data
- project\_grade\_category : categorical data
- teacher\_prefix : categorical data
- project\_title : text data
- text : text data
- project\_resource\_summary: text data (optinal)
- quantity : numerical (optinal)
- teacher\_number\_of\_previously\_posted\_projects : numerical
- price : numerical

## 2. Decision Tree

### 2.1 Splitting data into Train and cross validation(or test): Stratified Sampling

```
In [0]:
```

```
from sklearn.model_selection import train_test_split

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

#Checking if there are any values other than 0 and 1
project_data['project_is_approved'].unique()
```

```
#https://answers.dataiku.com/2352/split-dataset-by-stratified-sampling
df_train, df_test = train_test_split(project_data, test_size = 0.3, stratify=project_data['project_
is_approved'])
print(df_train.shape,df_test.shape)
```

```
(35000, 25) (15000, 25)
```

## 2.2 Make Data Model Ready: encoding numerical, categorical features

### 2.2.1 Vectorizing Categorical data

#### 2.2.1.1 Feature encoding for categories

In [0]:

```
# we use count vectorizer to convert the values into one
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer(vocabulary=list(sorted_cat_dict.keys()), lowercase=False, binary=True
)
categories_one_hot_train = vectorizer.fit_transform(df_train['clean_categories'].values)
categories_one_hot_test = vectorizer.transform(df_test['clean_categories'].values)
print(vectorizer.get_feature_names())
print("Shape of matrices after one hot encoding ",categories_one_hot_train.shape,
categories_one_hot_test.shape)

# Store feature names in a list
feature_names = []
feature_names = vectorizer.get_feature_names()
print(len(feature_names))
```

```
['Warmth', 'Care_Hunger', 'History_Civics', 'Music_Arts', 'AppliedLearning', 'SpecialNeeds',
'Health_Sports', 'Math_Science', 'Literacy_Language']
Shape of matrices after one hot encoding (35000, 9) (15000, 9)
9
```

#### 2.2.1.2 Feature encoding for subcategories

In [0]:

```
# we use count vectorizer to convert the values into one
vectorizer = CountVectorizer(vocabulary=list(sorted_sub_cat_dict.keys()), lowercase=False, binary=
True)
sub_categories_one_hot_train = vectorizer.fit_transform(df_train['clean_subcategories'].values)
sub_categories_one_hot_test = vectorizer.transform(df_test['clean_subcategories'].values)
print(vectorizer.get_feature_names())
print("Shape of matrices after one hot encoding ",sub_categories_one_hot_train.shape,
sub_categories_one_hot_test.shape)

# Store feature names in a list
feature_names.extend(vectorizer.get_feature_names())
print(len(feature_names))
```

```
['Economics', 'CommunityService', 'FinancialLiteracy', 'ParentInvolvement', 'Extracurricular',
'Civics_Government', 'ForeignLanguages', 'Warmth', 'Care_Hunger', 'NutritionEducation',
'SocialSciences', 'PerformingArts', 'CharacterEducation', 'TeamSports', 'Other',
'College_CareerPrep', 'Music', 'History_Geography', 'Health_LifeScience', 'EarlyDevelopment', 'ESL
', 'Gym_Fitness', 'EnvironmentalScience', 'VisualArts', 'Health_Wellness', 'AppliedSciences',
'SpecialNeeds', 'Literature_Writing', 'Mathematics', 'Literacy']
Shape of matrices after one hot encoding (35000, 30) (15000, 30)
39
```

#### 2.2.1.3 Feature encoding for state

In [0]:

```
# we use count vectorizer to convert the values into one hot encoded features

#https://cmdlinetips.com/2018/01/how-to-get-unique-values-from-a-column-in-pandas-data-frame/
#To get unique values from school_state column

school_state_lst=project_data['school_state'].unique()

vectorizer = CountVectorizer(vocabulary = school_state_lst, lowercase=False, binary=True)

school_state_one_hot_train = vectorizer.fit_transform(df_train['school_state'].values)
school_state_one_hot_test = vectorizer.transform(df_test['school_state'].values)
print(vectorizer.get_feature_names())
print("Shape of matrices after one hot encoding
",school_state_one_hot_train.shape,school_state_one_hot_test.shape)

# Store feature names in a list
feature_names.extend(vectorizer.get_feature_names())
print(len(feature_names))

['GA', 'CA', 'OH', 'FL', 'MD', 'TX', 'NJ', 'OK', 'PA', 'WV', 'NC', 'CO', 'VA', 'AZ', 'MA', 'ID', 'M
I', 'ME', 'WA', 'SC', 'LA', 'TN', 'MS', 'IN', 'KS', 'NY', 'KY', 'WI', 'MO', 'IA', 'SD', 'UT', 'IL',
'CT', 'NV', 'AL', 'MN', 'AR', 'DC', 'OR', 'NH', 'RI', 'HI', 'NE', 'NM', 'AK', 'ND', 'DE', 'MT', 'VT
', 'WY']
Shape of matrices after one hot encoding   (35000, 51) (15000, 51)
90
```

#### 2.2.1.4 Feature encoding for teacher\_prefix

In [0]:

```
# we use count vectorizer to convert the values into one hot encoded features

#https://cmdlinetips.com/2018/01/how-to-get-unique-values-from-a-column-in-pandas-data-frame/
#https://stackoverflow.com/questions/48090658/sklearn-how-to-incorporate-missing-data-when-one-hot
-encoding

#fetching unique values
teacher_prefix_lst=project_data['teacher_prefix'].unique()

vectorizer = CountVectorizer(vocabulary = teacher_prefix_lst, lowercase=False, binary=True)

teacher_prefix_one_hot_train = vectorizer.fit_transform(df_train['teacher_prefix'].values)
teacher_prefix_one_hot_test = vectorizer.transform(df_test['teacher_prefix'].values)
print(vectorizer.get_feature_names())
print("Shape of matrices after one hot encoding
",teacher_prefix_one_hot_train.shape,teacher_prefix_one_hot_test.shape)

# Store feature names in a list
feature_names.extend(vectorizer.get_feature_names())
print(len(feature_names))

['Mrs.', 'Ms.', 'Mr.', 'Teacher', 'Unknown', 'Dr.']
Shape of matrices after one hot encoding   (35000, 6) (15000, 6)
96
```

#### 2.2.1.5 Feature encoding for project\_grade\_category

In [0]:

```
# we use count vectorizer to convert the values into one hot encoded features

#https://cmdlinetips.com/2018/01/how-to-get-unique-values-from-a-column-in-pandas-data-frame/
#To get unique values from project_grade_category column
grade_cat_lst=project_data['project_grade_category'].unique()

vectorizer = CountVectorizer(vocabulary = grade_cat_lst, lowercase=False, binary=True)

grade_cat_one_hot_train = vectorizer.fit_transform(df_train['project_grade_category'].values)
grade_cat_one_hot_test = vectorizer.transform(df_test['project_grade_category'].values)
print(vectorizer.get_feature_names())
print("Shape of matrix after one hot encoding ",grade_cat_one_hot_train.shape,
```

```

grade_cat_one_hot_test.shape)

# Store feature names in a list
feature_names.extend(vectorizer.get_feature_names())
print(len(feature_names))

['Grades PreK-2', 'Grades 3-5', 'Grades 6-8', 'Grades 9-12']
Shape of matrix after one hot encoding (35000, 4) (15000, 4)
100

```

## 2.2.2 Vectorizing Numerical features

### 2.2.2.1 Vectorizing price

In [0]:

```

# check this one: https://www.youtube.com/watch?v=0HOqOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler

# Reshape your data either using array.reshape(-1, 1)
print(df_train.columns)
price_scaler = StandardScaler()
price_scaler.fit(df_train['price'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {price_scaler.mean_[0]}, Standard deviation : {np.sqrt(price_scaler.var_[0])}")

# Now standardize the data with above mean and variance.
price_train_standardized = price_scaler.transform(df_train['price'].values.reshape(-1, 1))
price_test_standardized = price_scaler.transform(df_test['price'].values.reshape(-1, 1))

# Store feature names in a list
feature_names.append('price')
print(len(feature_names))

```

```

Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects', 'project_is_approved',
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc',
      'senti_score'],
      dtype='object')
Mean : 311.2684708571428, Standard deviation : 380.1424180619439
101

```

### 2.2.2.2 Vectorizing no. of previously posted projects

In [0]:

```

# check this one: https://www.youtube.com/watch?v=0HOqOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings("ignore")

prev_proj_scaler = StandardScaler()
prev_proj_scaler.fit(df_train['teacher_number_of_previously_posted_projects'].values.reshape(-1,1))
# finding the mean and standard deviation of this data
print(f"Mean : {prev_proj_scaler.mean_[0]}, Standard deviation : {np.sqrt(prev_proj_scaler.var_[0])}")

# Now standardize the data with above mean and variance.
prev_proj_train_standardized =
prev_proj_scaler.transform(df_train['teacher_number_of_previously_posted_projects'].values.reshape(-1, 1))
prev_proj_test_standardized =

```

```
prev_proj_scalar.transform(df_test['teacher_number_of_previously_posted_projects'].values.reshape(-1, 1))

# Store feature names in a list
feature_names.append('teacher_number_of_previously_posted_projects')
print(len(feature_names))
```

Mean : 10.185285714285714, Standard deviation : 25.750309524987326  
102

### 2.2.2.3 Vectorizing word counts of project title

In [0]:

```
# check this one: https://www.youtube.com/watch?v=0HOqOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings("ignore")

wc_title_scalar = StandardScaler()
wc_title_scalar.fit(df_train['title_wc'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {wc_title_scalar.mean_[0]}, Standard deviation : {np.sqrt(wc_title_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
wc_title_train_standardized = wc_title_scalar.transform(df_train['title_wc'].values.reshape(-1, 1))
wc_title_test_standardized = wc_title_scalar.transform(df_test['title_wc'].values.reshape(-1, 1))

# Store feature names in a list
feature_names.append('title_wc')
print(len(feature_names))
```

Mean : 3.6558285714285716, Standard deviation : 1.5350208092719653  
103

### 2.2.2.4 Vectorizing word counts of essay text

In [0]:

```
# check this one: https://www.youtube.com/watch?v=0HOqOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings("ignore")

wc_essay_scalar = StandardScaler()
wc_essay_scalar.fit(df_train['essay_wc'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {wc_essay_scalar.mean_[0]}, Standard deviation : {np.sqrt(wc_essay_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
wc_essay_train_standardized = wc_essay_scalar.transform(df_train['essay_wc'].values.reshape(-1, 1))
wc_essay_test_standardized = wc_essay_scalar.transform(df_test['essay_wc'].values.reshape(-1, 1))

# Store feature names in a list
feature_names.append('essay_wc')
print(len(feature_names))
```

Mean : 136.67754285714287, Standard deviation : 35.67317309315051  
104

### 2.2.2.5 Vectorizing sentimental scores of project essays

In [0]:



In [0]:

```
# check this one: https://www.youtube.com/watch?v=0HOqOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings("ignore")

senti_score_scalar = StandardScaler()
senti_score_scalar.fit(df_train['senti_score'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {senti_score_scalar.mean_[0]}, Standard deviation : {np.sqrt(senti_score_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
senti_score_train_standardized =
senti_score_scalar.transform(df_train['senti_score'].values.reshape(-1, 1))
senti_score_test_standardized = senti_score_scalar.transform(df_test['senti_score'].values.reshape(-1, 1))

# Store feature names in a list
feature_names.append('senti_score')
print(len(feature_names))
```

Mean : 0.9590257657142857, Standard deviation : 0.15127913990127925  
105

### 2.2.2.6 Vectorizing Quantity

In [0]:

```
# check this one: https://www.youtube.com/watch?v=0HOqOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings("ignore")

qty_scalar = StandardScaler()
qty_scalar.fit(df_train['quantity'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {qty_scalar.mean_[0]}, Standard deviation : {np.sqrt(qty_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
qty_train_standardized = qty_scalar.transform(df_train['quantity'].values.reshape(-1, 1))
qty_test_standardized = qty_scalar.transform(df_test['quantity'].values.reshape(-1, 1))

# Store feature names in a list
feature_names.append('quantity')
print(len(feature_names))

# Store feature names in a list
feature_names_tfidf = feature_names.copy()
feature_names_bow = feature_names.copy()
print(len(feature_names_bow), len(feature_names_tfidf))
```

Mean : 17.676, Standard deviation : 27.40447713161378  
106  
106 106

## 2.3 Make Data Model Ready: encoding eassay, and project\_title

### 2.3.1 Vectorizing Text data

#### 2.3.1.1 Bag of words for essay text

In [0]:

```
# We are considering only the words which appeared in at least 10 documents(rows or projects).
vectorizer = CountVectorizer(min_df=10)
text_train_bow = vectorizer.fit_transform(df_train['clean_essay'])
text_test_bow = vectorizer.transform(df_test['clean_essay'])
print("Shape of matrix after one hot encoding ",text_train_bow.shape, text_test_bow.shape)

# Store feature names in a list
feature_names_bow.extend(vectorizer.get_feature_names())
print(len(feature_names_bow))
```

Shape of matrix after one hot encoding (35000, 10491) (15000, 10491)  
10597

In [0]:

```
# you can vectorize the title also
# before you vectorize the title make sure you preprocess it

vectorizer = CountVectorizer(min_df=10)
title_train_bow = vectorizer.fit_transform(df_train['clean_title'])
title_test_bow = vectorizer.transform(df_test['clean_title'])
print("Shape of matrix after one hot encoding ", title_train_bow.shape, title_test_bow.shape)

# Store feature names in a list
feature_names_bow.extend(vectorizer.get_feature_names())
print(len(feature_names_bow))
```

Shape of matrix after one hot encoding (35000, 1579) (15000, 1579)  
12176

### 2.3.1.2 TFIDF vectorizer for essay text

In [0]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(min_df=10)

text_train_tfidf = vectorizer.fit_transform(df_train['clean_essay'])
text_test_tfidf = vectorizer.transform(df_test['clean_essay'])
print("Shape of matrix after one hot encoding ",text_train_tfidf.shape, text_test_tfidf.shape)

# Store feature names in a list
feature_names_tfidf.extend(vectorizer.get_feature_names())
print(len(feature_names_tfidf))
```

Shape of matrix after one hot encoding (35000, 10491) (15000, 10491)  
10597

In [0]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(min_df=10)

title_train_tfidf = vectorizer.fit_transform(df_train['clean_title'])
title_test_tfidf = vectorizer.transform(df_test['clean_title'])

print("Shape of matrix after one hot encoding ",title_train_tfidf.shape, title_test_tfidf.shape)

# Store feature names in a list
feature_names_tfidf.extend(vectorizer.get_feature_names())
print(len(feature_names_tfidf))
```

Shape of matrix after one hot encoding (35000, 1579) (15000, 1579)  
12176

### 2.3.1.3 Using Pretrained models: Avg W2V vectorizer

In [0]:

```
'''def loadGloveModel(gloveFile):
    print ("Loading Glove Model")
    f = open(gloveFile,'r', encoding="utf8")
    model = {}
    for line in tqdm(f):
        splitLine = line.split()
        word = splitLine[0]
        embedding = np.array([float(val) for val in splitLine[1:]])
        model[word] = embedding
    print ("Done.",len(model)," words loaded!")
    return model
model = loadGloveModel('drive/My Drive/Colab Notebooks/glove.6B.50d.txt')'''
```

Out[0]:

```
'def loadGloveModel(gloveFile):\n    print ("Loading Glove Model")\n    f = open(gloveFile,\'r\',\nencoding="utf8")\n    model = {}\n    for line in tqdm(f):\n        splitLine = line.split()\nword = splitLine[0]\n        embedding = np.array([float(val) for val in splitLine[1:]])\n    model[word] = embedding\n    print ("Done.",len(model)," words loaded!")\n    return model\nmodel =\nloadGloveModel(\'drive/My Drive/Colab Notebooks/glove.6B.50d.txt\')
```

In [0]:

```
'''words = []
for i in preprocessed_essays:
    words.extend(i.split(' '))

for i in preprocessed_title:
    words.extend(i.split(' '))
print("all the words in the coupus", len(words))
words = set(words)
print("the unique words in the coupus", len(words))

inter_words = set(model.keys()).intersection(words)
print("The number of words that are present in both glove vectors and our coupus", \
      len(inter_words), "("",np.round(len(inter_words)/len(words)*100,3),"%)")

words_courpus = {}
words_glove = set(model.keys())
for i in words:
    if i in words_glove:
        words_courpus[i] = model[i]
print("word 2 vec length", len(words_courpus))

# stronging variables into pickle files python: http://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-python/

import pickle
with open('drive/My Drive/Colab Notebooks/glove_vectors_50', 'wb') as f:
    pickle.dump(words_courpus, f)'''
```

Out[0]:

```
'words = []\nfor i in preprocessed_essays:\n    words.extend(i.split(\' \''))\n\nfor i in\npreprocessed_title:\n    words.extend(i.split(\' \''))\n\nprint("all the words in the coupus",\nlen(words))\n\nwords = set(words)\n\nprint("the unique words in the coupus",\nlen(words))\n\n\ninter_words = set(model.keys()).intersection(words)\n\nprint("The number of words tha\nt are present in both glove vectors and our coupus",\n      len(inter_words),\n      "("",np.round(len(inter_words)/len(words)*100,3),"%)")\n\n\nwords_courpus = {}\nwords_glove =\nset(model.keys())\n\nfor i in words:\n    if i in words_glove:\n        words_courpus[i] = model[i]\n\nprint("word 2 vec length", len(words_courpus))\n\n\n# stronging variables into pickle files python\n: http://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-python/\n\n\nimport pic\nkle\n\nwith open(\'drive/My Drive/Colab Notebooks/glove_vectors_50\', \'wb\') as f:\n\npickle.dump(words_courpus, f)'
```

In [0]:

```
# storing variables into pickle files python: http://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-python/
# make sure you have the glove_vectors file

with open('drive/My Drive/Colab Notebooks/glove_vectors_50', 'rb') as f:
    model = pickle.load(f)
```

```
model = pickle.load(f)
glove_words = set(model.keys())
```

In [0]:

```
# average Word2Vec
# compute average word2vec for each review.

avg_w2v_train_text_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_train['clean_essay']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    cnt_words = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if word in glove_words:
            vector += model[word]
            cnt_words += 1
    if cnt_words != 0:
        vector /= cnt_words
    avg_w2v_train_text_vectors.append(vector)

print(len(avg_w2v_train_text_vectors))
print(len(avg_w2v_train_text_vectors[0]))
```

100%|██████████| 35000/35000 [00:06<00:00, 5183.68it/s]

35000  
50

In [0]:

```
# average Word2Vec
# compute average word2vec for each review.

avg_w2v_test_text_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_test['clean_essay']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    cnt_words = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if word in glove_words:
            vector += model[word]
            cnt_words += 1
    if cnt_words != 0:
        vector /= cnt_words
    avg_w2v_test_text_vectors.append(vector)

print(len(avg_w2v_test_text_vectors))
print(len(avg_w2v_test_text_vectors[0]))
```

100%|██████████| 15000/15000 [00:02<00:00, 5232.91it/s]

15000  
50

In [0]:

```
# Similarly you can vectorize for title also

# average Word2Vec
# compute average word2vec for each title
avg_w2v_title_train_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_train['clean_title']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    cnt_words = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if word in glove_words:
            vector += model[word]
```

```

        vector += model[word]
        cnt_words += 1
    if cnt_words != 0:
        vector /= cnt_words
    avg_w2v_title_train_vectors.append(vector)

print(len(avg_w2v_title_train_vectors))
print(len(avg_w2v_title_train_vectors[0]))

```

100%|██████████| 35000/35000 [00:00<00:00, 107766.21it/s]

35000

50

In [0]:

```

# Similarly you can vectorize for title also

# average Word2Vec
# compute average word2vec for each title
avg_w2v_title_test_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_test['clean_title']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    cnt_words = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if word in glove_words:
            vector += model[word]
            cnt_words += 1
    if cnt_words != 0:
        vector /= cnt_words
    avg_w2v_title_test_vectors.append(vector)

print(len(avg_w2v_title_test_vectors))
print(len(avg_w2v_title_test_vectors[0]))

```

100%|██████████| 15000/15000 [00:00<00:00, 105148.34it/s]

15000

50

#### 2.3.1.4 Using Pretrained Models: TFIDF weighted W2V for essay text

In [0]:

```

# S = ["abc def pqr", "def def def abc", "pqr pqr def"]
tfidf_model = TfidfVectorizer()
tfidf_model.fit_transform(df_train['clean_essay'])
# we are converting a dictionary with word as a key, and the idf as a value
dictionary = dict(zip(tfidf_model.get_feature_names(), list(tfidf_model.idf_)))
tfidf_words = set(tfidf_model.get_feature_names())

```

In [0]:

```

# average Word2Vec
# compute average word2vec for each review.
tfidf_w2v_train_text_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_train['clean_essay']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    tf_idf_weight = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if (word in glove_words) and (word in tfidf_words):
            vec = model[word] # getting the vector for each word
            # here we are multiplying idf value(dictionary[word]) and the tf value((sentence.count(word)/len(sentence.split())))
            tf_idf = dictionary[word]*(sentence.count(word)/len(sentence.split())) # getting the tf idf value for each word
            vector += tf_idf * vec
    if tf_idf_weight != 0:
        vector /= tf_idf_weight
    tfidf_w2v_train_text_vectors.append(vector)

```

```

for value in each word:
    vector += (vec * tf_idf) # calculating tfidf weighted w2v
    tf_idf_weight += tf_idf
    if tf_idf_weight != 0:
        vector /= tf_idf_weight
    tfidf_w2v_train_text_vectors.append(vector)

print(len(tfidf_w2v_train_text_vectors))
print(len(tfidf_w2v_train_text_vectors[0]))

```

100%|██████████| 35000/35000 [00:47<00:00, 731.66it/s]

35000  
50

In [0]:

```

# average Word2Vec
# compute average word2vec for each review.
tfidf_w2v_test_text_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_test['clean_essay']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    tf_idf_weight = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if (word in glove_words) and (word in tfidf_words):
            vec = model[word] # getting the vector for each word
            # here we are multiplying idf value(dictionary[word]) and the tf
            value((sentence.count(word)/len(sentence.split())))
            tf_idf = dictionary[word]*(sentence.count(word)/len(sentence.split())) # getting the tf
            idf value for each word
            vector += (vec * tf_idf) # calculating tfidf weighted w2v
            tf_idf_weight += tf_idf
    if tf_idf_weight != 0:
        vector /= tf_idf_weight
    tfidf_w2v_test_text_vectors.append(vector)

print(len(tfidf_w2v_test_text_vectors))
print(len(tfidf_w2v_test_text_vectors[0]))

```

100%|██████████| 15000/15000 [00:20<00:00, 729.45it/s]

15000  
50

#### 2.3.1.4 Using Pretrained Models: TFIDF weighted W2V for title

In [0]:

```

# Similarly you can vectorize for title also

tfidf_model = TfidfVectorizer()
tfidf_model.fit_transform(df_train['clean_title'])
# we are converting a dictionary with word as a key, and the idf as a value
dictionary = dict(zip(tfidf_model.get_feature_names(), list(tfidf_model.idf_)))
tfidf_words = set(tfidf_model.get_feature_names())

```

In [0]:

```

# average Word2Vec
# compute average word2vec for each project title.
tfidf_w2v_train_title_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_train['clean_title']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    tf_idf_weight = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if (word in glove_words) and (word in tfidf_words):
            vec = model[word] # getting the vector for each word
            # here we are multiplying idf value(dictionary[word]) and the tf
            value((sentence.count(word)/len(sentence.split())))
            tf_idf = dictionary[word]*(sentence.count(word)/len(sentence.split())) # getting the tf
            idf value for each word
            vector += (vec * tf_idf) # calculating tfidf weighted w2v
            tf_idf_weight += tf_idf
    if tf_idf_weight != 0:
        vector /= tf_idf_weight
    tfidf_w2v_train_title_vectors.append(vector)

```

```

        if (word in glove_words) and (word in tfidf_words):
            vec = model[word] # getting the vector for each word
            # here we are multiplying idf value(dictionary[word]) and the tf
            value((sentence.count(word)/len(sentence.split())))
            tf_idf = dictionary[word]*(sentence.count(word)/len(sentence.split())) # getting the tf
            idf value for each word
            vector += (vec * tf_idf) # calculating tfidf weighted w2v
            tf_idf_weight += tf_idf
        if tf_idf_weight != 0:
            vector /= tf_idf_weight
        tfidf_w2v_train_title_vectors.append(vector)

print(len(tfidf_w2v_train_title_vectors))
print(len(tfidf_w2v_train_title_vectors[0]))

```

100%|██████████| 35000/35000 [00:00<00:00, 51853.16it/s]

35000  
50

In [0]:

```

# average Word2Vec
# compute average word2vec for each project title.
tfidf_w2v_test_title_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(df_test['clean_title']): # for each review/sentence
    vector = np.zeros(50) # as word vectors are of zero length. 50 is the size of each vector in glove file
    tf_idf_weight = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if (word in glove_words) and (word in tfidf_words):
            vec = model[word] # getting the vector for each word
            # here we are multiplying idf value(dictionary[word]) and the tf
            value((sentence.count(word)/len(sentence.split())))
            tf_idf = dictionary[word]*(sentence.count(word)/len(sentence.split())) # getting the tf
            idf value for each word
            vector += (vec * tf_idf) # calculating tfidf weighted w2v
            tf_idf_weight += tf_idf
        if tf_idf_weight != 0:
            vector /= tf_idf_weight
        tfidf_w2v_test_title_vectors.append(vector)

print(len(tfidf_w2v_test_title_vectors))
print(len(tfidf_w2v_test_title_vectors[0]))

```

100%|██████████| 15000/15000 [00:00<00:00, 51553.80it/s]

15000  
50

## 2.4 Applying Decision Tree Classifier on different kinds of featurizations as mentioned in the instructions

### 2.4.1 Applying Decision Tree Classifier on BOW featurization, SET 1

#### Hyper paramter tuning method: GridSearch

In [0]:

```

#https://www.digitalocean.com/community/tutorials/how-to-plot-data-in-python-3-using-matplotlib
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html
#https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter

from scipy.sparse import hstack
from sklearn.model_selection import GridSearchCV

```

```
import matplotlib.patches as mpatches
from sklearn.metrics import roc_auc_score

print(type(categories_one_hot_train), type(sub_categories_one_hot_train),
type(grade_cat_one_hot_train),
      type(teacher_prefix_one_hot_train), type(school_state_one_hot_train), type(price_
train_standardized),
      type(prev_proj_train_standardized), type(wc_title_train_standardized), type(wc_es
say_train_standardized), type(senti_score_train_standardized),
      type(qty_train_standardized), type(text_train_bow), type(title_train_bow))

x_train = hstack((categories_one_hot_train, sub_categories_one_hot_train, grade_cat_one_hot_train,
teacher_prefix_one_hot_train, school_state_one_hot_train,
price_train_standardized,
prev_proj_train_standardized, wc_title_train_standardized,
wc_essay_train_standardized, senti_score_train_standardized,
qty_train_standardized, text_train_bow, title_train_bow))
y_train = df_train['project_is_approved']

x_test = hstack((categories_one_hot_test, sub_categories_one_hot_test, grade_cat_one_hot_test,
teacher_prefix_one_hot_test, school_state_one_hot_test, price_test_standardized,
prev_proj_test_standardized, wc_title_test_standardized,
wc_essay_test_standardized, senti_score_test_standardized,
qty_test_standardized, text_test_bow, title_test_bow))
y_test = df_test['project_is_approved']

print(x_train.shape, type(x_train), y_train.shape, type(y_train))
print(x_test.shape, type(x_test), y_test.shape, type(y_test))
```

```
<class 'scipy.sparse.csr.csr_matrix'> <class 'scipy.sparse.csr.csr_matrix'> <class  
'scipy.sparse.csr.csr_matrix'> <class 'scipy.sparse.csr.csr_matrix'> <class  
'scipy.sparse.csr.csr_matrix'> <class 'numpy.ndarray'> <class 'numpy.ndarray'> <class  
'numpy.ndarray'> <class 'numpy.ndarray'> <class 'numpy.ndarray'> <class 'numpy.ndarray'> <class 's  
cipy.sparse.csr.csr_matrix'> <class 'scipy.sparse.csr.csr_matrix'>  
(35000, 12176) <class 'scipy.sparse.coo.coo_matrix'> (35000,) <class 'pandas.core.series.Series'>  
(15000, 12176) <class 'scipy.sparse.coo.coo_matrix'> (15000,) <class 'pandas.core.series.Series'>
```

In [0]:

```
import dill
#dill.dump_session('sess_knn.pckl')
#dill.load_session('sess_knn.pckl')

print(x_train.todense()[1])
```

```
[[0. 0. 0. ... 0. 0. 0.]]
```

In [0]:

```
#https://stackabuse.com/cross-validation-and-grid-search-for-model-selection-in-python/
#https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

#Initialising Classifier
classifier = DecisionTreeClassifier(class_weight='balanced')

#Brute force approach for finding best K value
parameters = {'max_depth': [1, 5, 10, 50, 100, 500, 100],
              'min_samples_split': [5, 10, 100, 500]}

#Training the model on train data
DT_BoW = GridSearchCV(classifier, parameters, cv=10, return_train_score=True, scoring='roc_auc', n_jobs=-1)
DT_BoW.fit(x_train, y_train)
```

Out[0]:

[illegible]



```

min_impurity_split=None,
min_samples_leaf=1,
min_samples_split=2,
min_weight_fraction_leaf=0.0,
presort=False, random_state=None,
splitter='best'),

iid='warn', n_jobs=-1,
param_grid={'max_depth': [1, 5, 10, 50, 100, 500, 1000],
            'min_samples_split': [5, 10, 100, 500]},
pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
scoring='roc_auc', verbose=0)

```

In [0]:

```

#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://stackoverflow.com/questions/20944483/python-3-sort-a-dict-by-its-values/20948781

```

```

print(DT_BoW.best_params_) #Gives the best value of parameters from the given range
print(DT_BoW.cv_results_['params'])

```

```

params_train = {}
params_test = {}

```

```

for i,j in zip(DT_BoW.cv_results_['params'],DT_BoW.cv_results_['mean_train_score']):
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))
    params_train[a]=j

```

```

for i,j in zip(DT_BoW.cv_results_['params'],DT_BoW.cv_results_['mean_test_score']):
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))
    params_test[a]=j

```

```

params_train = {k: params_train[k] for k in sorted(params_train)}
params_test = {k: params_test[k] for k in sorted(params_test)}

```

```

print(params_train)
print(params_test)

```

```

plt.figure(figsize=(20,3))
plt.plot(params_train.keys(),params_train.values(), label="Train")
plt.plot(params_test.keys(),params_test.values(), label="Test")
plt.title('AUC plot for ')
plt.xlabel('Parameters')
plt.ylabel('Area under ROC Curve')
plt.legend()
plt.grid()
plt.show()
plt.close()

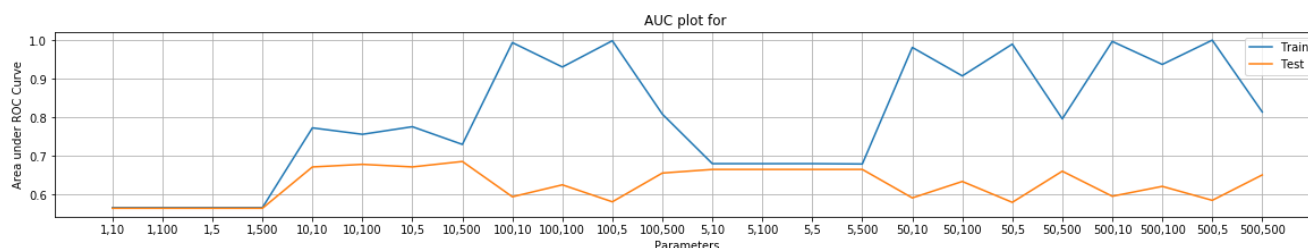
```

```

{'max_depth': 10, 'min_samples_split': 500}
[{'max_depth': 1, 'min_samples_split': 5}, {'max_depth': 1, 'min_samples_split': 10},
{'max_depth': 1, 'min_samples_split': 100}, {'max_depth': 1, 'min_samples_split': 500},
{'max_depth': 5, 'min_samples_split': 5}, {'max_depth': 5, 'min_samples_split': 10}, {'max_depth': 5, 'min_samples_split': 100}, {'max_depth': 5, 'min_samples_split': 500}, {'max_depth': 10, 'min_samples_split': 5}, {'max_depth': 10, 'min_samples_split': 10}, {'max_depth': 10, 'min_samples_split': 100}, {'max_depth': 10, 'min_samples_split': 500}, {'max_depth': 50, 'min_samples_split': 5}, {'max_depth': 50, 'min_samples_split': 10}, {'max_depth': 50, 'min_samples_split': 100}, {'max_depth': 50, 'min_samples_split': 500}, {'max_depth': 100, 'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100, 'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}, {'max_depth': 500, 'min_samples_split': 5}, {'max_depth': 500, 'min_samples_split': 10}, {'max_depth': 500, 'min_samples_split': 100}, {'max_depth': 500, 'min_samples_split': 500}, {'max_depth': 100, 'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100, 'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}]
{'1,10': 0.5658447630496513, '1,100': 0.5658447630496513, '1,5': 0.5658447630496513, '1,500': 0.5658447630496513, '10,10': 0.7725327779649365, '10,100': 0.7558363880297121, '10,5': 0.7755068322923123, '10,500': 0.7296227215982557, '100,10': 0.9934954527905304, '100,100': 0.9302552266097175, '100,5': 0.9983128695325562, '100,500': 0.8082024368287503, '5,10': 0.6797779097850962, '5,100': 0.6797194323611472, '5,5': 0.679787391527868, '5,500': 0.6789459883030039, '50,10': 0.9810245090955639, '50,100': 0.9071937046479871, '50,5': 0.9895632800758541, '50,500': 0.7959066819223961, '500,10': 0.9965075944363987, '500,100': 0.9369270099488013, '500,5': 0.9997969006138803, '500,500': 0.8136928231228634}
{'1,10': 0.5641422578751661, '1,100': 0.5641422578751661, '1,5': 0.5641422578751661, '1,500': 0.5641422578751661, '10,10': 0.6712244788405152, '10,100': 0.6780032563310472, '10,5': 0.6713129852899227, '10,500': 0.68541390388202, '100,10': 0.5939473316373243, '100,100':

```

```
0.071312300203227, 10,500 : 0.00041300000202, 100,10 : 0.0003273310373243, 100,100 :
0.624903389127523, '100,5': 0.5809169738281615, '100,500': 0.655446492348402, '5,10':
0.6647949170096533, '5,100': 0.6648311997138651, '5,5': 0.6647687742888339, '5,500':
0.6648436191655691, '50,10': 0.5910526079561643, '50,100': 0.6336166733770557, '50,5':
0.5795304156286145, '50,500': 0.6602438585327474, '500,10': 0.5953351571734882, '500,100':
0.6209352906147195, '500,5': 0.584715449378995, '500,500': 0.6505672207884442}
```



In [0]:

```
#https://datascience.stackexchange.com/questions/22762/understanding-predict-proba-from-
multioutputclassifier
#https://stackoverflow.com/questions/34894587/should-we-plot-the-roc-curve-for-each-class

from sklearn.metrics import roc_curve, auc

#training the model on the best K value found in the above result
final_DT_BoW = DecisionTreeClassifier(max_depth=10, min_samples_split=500, class_weight='balanced'
)
final_DT_BoW.fit(x_train,y_train)

x_train_csr=x_train.tocsr()
x_test_csr=x_test.tocsr()

y_train_pred=[]
y_test_pred=[]

#ROC curve function takes the actual values and the predicted probabilities of the positive class
for i in range(0,x_train.shape[0]):
    y_train_pred.extend(final_DT_BoW.predict_proba(x_train_csr[i])[:,1]) #[:,1] gives the probabili
ty for class 1

for i in range(0,x_test.shape[0]):
    y_test_pred.extend(final_DT_BoW.predict_proba(x_test_csr[i])[:,1])
```

In [0]:

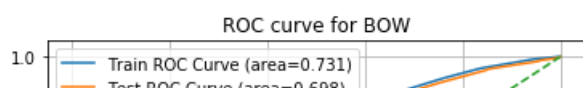
```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

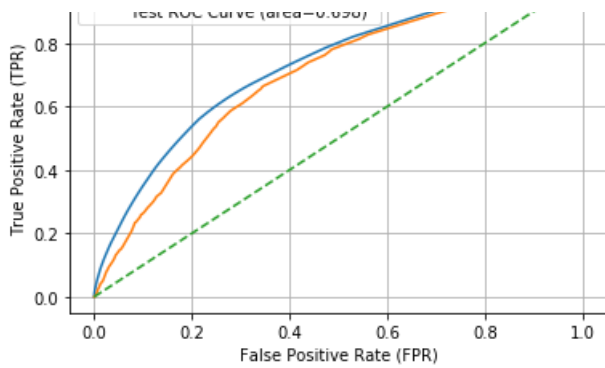
#Calculating FPR and TPR for train and test data
train_fpr, train_tpr, train_thresholds = roc_curve(y_train, y_train_pred)
test_fpr, test_tpr, test_thresholds = roc_curve(y_test, y_test_pred)

#Calculating AUC for train and test curves
roc_auc_train=auc(train_fpr,train_tpr)
roc_auc_test=auc(test_fpr,test_tpr)

plt.plot(train_fpr, train_tpr, label="Train ROC Curve (area=%0.3f)" % roc_auc_train)
plt.plot(test_fpr, test_tpr, label="Test ROC Curve (area=%0.3f)" % roc_auc_test)
plt.plot([0,1],[0,1],linestyle='--')
plt.legend()
plt.xlabel("False Positive Rate (FPR)")
plt.ylabel("True Positive Rate (TPR)")
plt.title("ROC curve for BOW")
plt.grid()

plt.show()
plt.close()
```





In [0]:

```
#https://stackoverflow.com/questions/35572000/how-can-i-plot-a-confusion-matrix
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
#https://datatofish.com/confusion-matrix-python/
```

```
from sklearn.metrics import confusion_matrix as cf_mx
```

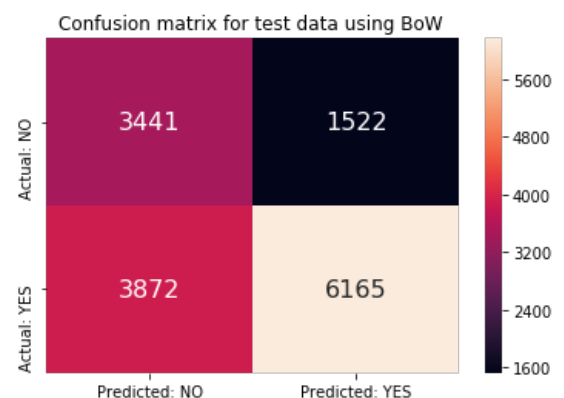
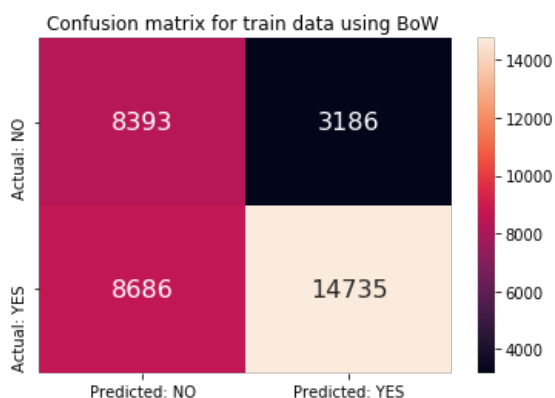
```
expected_train = y_train.values
predicted_train = final_DT_BoW.predict(x_train)
```

```
expected_test = y_test.values
predicted_test = final_DT_BoW.predict(x_test)
```

In [0]:

```
plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_train, predicted_train)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for train data using BoW ')
```

```
plt.subplot(1,2,2)
cmdf_test=cf_mx(expected_test, predicted_test)
df_cm_test = pd.DataFrame(cmdf_test, range(2),range(2))
df_cm_test.columns = ['Predicted: NO','Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for test data using BoW ')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()
```



## Decision Tree visualisation using Graphviz

In [0]:

```
#https://scikit-learn.org/stable/modules/generated/sklearn.tree.export_graphviz.html
```

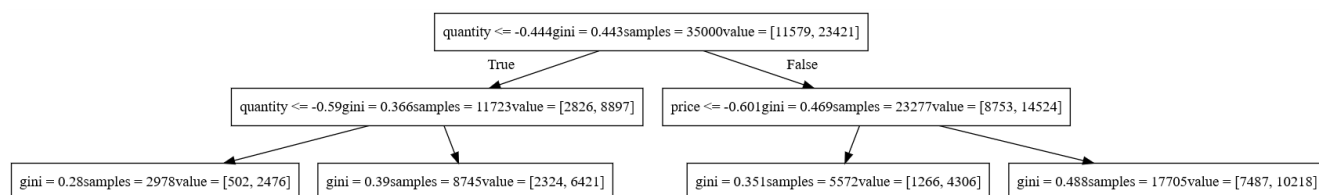
```
from sklearn import tree

clf = tree.DecisionTreeClassifier(max_depth=2)

clf = clf.fit(x_train, y_train)
tree.export_graphviz(clf, feature_names=feature_names_bow)
```

Out[0]:

```
'digraph Tree {\nnode [shape=box] ;\n0 [label="quantity <= -0.444\\ngini = 0.443\\nsamples = 35000\n\\nvalue = [11579, 23421]" ] ;\n1 [label="quantity <= -0.59\\ngini = 0.366\\nsamples =\n11723\\nvalue = [2826, 8897]" ] ;\n0 -> 1 [labeldistance=2.5, labelangle=45, headlabel="True"] ;\n2 [label="gini = 0.28\\nsamples = 2978\\nvalue = [502, 2476]" ] ;\n1 -> 2 ;\n3 [label="gini =\n0.39\\nsamples = 8745\\nvalue = [2324, 6421]" ] ;\n1 -> 3 ;\n4 [label="price <= -0.601\\ngini = 0.4\n69\\nsamples = 23277\\nvalue = [8753, 14524]" ] ;\n0 -> 4 [labeldistance=2.5, labelangle=-45, headl\nabel="False"] ;\n5 [label="gini = 0.351\\nsamples = 5572\\nvalue = [1266, 4306]" ] ;\n4 -> 5 ;\n6 [label="gini = 0.488\\nsamples = 17705\\nvalue = [7487, 10218]" ] ;\n4 -> 6 ;\n}
```



## Word cloud for False Positives

In [0]:

```
#https://stackoverflow.com/questions/36184432/is-it-possible-to-retrieve-false-positives-false-neg
atives-identified-by-a-conf
#https://stackoverflow.com/questions/31324218/scikit-learn-how-to-obtain-true-positive-true-negati
ve-false-positive-and-fal?rq=1

print(len(y_test), len(predicted_test))

index=[]

y_test_fp = y_test.values.reshape(-1,1)

#Comparing each value of actual y value and predicted y value, in D_Test
#False positive: The actual value is negative(0), but the model predicted it to be positive(1)
for i in range(len(y_test)):
    if (y_test_fp[i] == 0) and (predicted_test[i] == 1):
        index.append(i) #Storing the index of the FP datapoints in a list

print(len(index))
print(index)
```

```
15000 15000
1522
[19, 22, 27, 32, 35, 41, 45, 70, 80, 87, 119, 123, 168, 185, 192, 203, 222, 232, 233, 235, 250, 29
9, 304, 320, 331, 351, 353, 382, 386, 388, 401, 414, 415, 418, 432, 436, 443, 444, 454, 462, 465,
488, 495, 511, 525, 528, 535, 537, 561, 563, 565, 566, 583, 586, 601, 603, 604, 608, 648, 672, 676
, 678, 682, 696, 713, 731, 736, 743, 754, 767, 774, 783, 789, 792, 796, 801, 809, 821, 823, 843, 8
49, 857, 860, 863, 865, 872, 892, 896, 915, 944, 950, 967, 986, 1003, 1004, 1022, 1048, 1062,
1064, 1071, 1084, 1091, 1098, 1103, 1105, 1144, 1150, 1151, 1152, 1158, 1206, 1208, 1212, 1234, 124
3, 1248, 1249, 1250, 1251, 1254, 1258, 1267, 1275, 1277, 1313, 1326, 1332, 1340, 1341, 1349, 1358,
1369, 1383, 1384, 1395, 1397, 1407, 1453, 1493, 1503, 1507, 1526, 1535, 1541, 1542, 1555, 1557, 156
2, 1563, 1574, 1576, 1606, 1610, 1626, 1634, 1637, 1640, 1659, 1664, 1672, 1676, 1685, 1686, 1706,
1731, 1743, 1761, 1767, 1793, 1797, 1812, 1820, 1838, 1841, 1847, 1848, 1851, 1864, 1879, 1885, 189
0, 1906, 1916, 1917, 1920, 1922, 1935, 1939, 1993, 2007, 2009, 2016, 2017, 2026, 2037, 2061, 2062,
2068, 2081, 2107, 2108, 2117, 2123, 2126, 2127, 2128, 2131, 2151, 2153, 2164, 2207, 2211, 2215, 221
6, 2226, 2231, 2259, 2269, 2270, 2272, 2274, 2285, 2317, 2322, 2333, 2339, 2346, 2358, 2371, 2373,
2380, 2397, 2422, 2428, 2431, 2436, 2442, 2453, 2476, 2478, 2481, 2490, 2494, 2505, 2509, 2518, 252
4, 2526, 2546, 2554, 2567, 2578, 2580, 2582, 2583, 2615, 2626, 2633, 2653, 2656, 2663, 2666, 2670,
2672, 2681, 2685, 2686, 2711, 2712, 2720, 2737, 2747, 2748, 2762, 2767, 2791, 2794, 2801, 2811, 282
5, 2839, 2844, 2859, 2862, 2885, 2886, 2927, 2931, 2953, 2957, 2980, 2994, 3000, 3007, 3014, 3020,
3027, 3028, 3051, 3064, 3083, 3084, 3089, 3097, 3108, 3114, 3119, 3120, 3126, 3138, 3142, 3149, 316
0, 3202, 3203, 3208, 3213, 3224, 3228, 3231, 3248, 3250, 3269, 3272, 3283, 3304, 3308, 3314, 3334,
3344, 3350, 3367, 3386, 3410, 3427, 3428, 3442, 3446, 3455, 3456, 3457, 3471, 3475, 3483, 3507, 351
```

3, 3524, 3526, 3527, 3538, 3545, 3555, 3559, 3568, 3580, 3581, 3597, 3614, 3616, 3617, 3628, 3645, 3657, 3660, 3672, 3682, 3687, 3723, 3733, 3748, 3755, 3760, 3768, 3774, 3775, 3781, 3789, 3793, 3795, 3805, 3807, 3818, 3819, 3821, 3822, 3824, 3829, 3832, 3839, 3844, 3861, 3864, 3910, 3926, 3941, 3945, 3946, 3961, 3972, 3973, 3975, 3981, 3994, 3996, 4002, 4003, 4004, 4015, 4019, 4023, 4028, 4029, 4030, 4039, 4056, 4076, 4107, 4112, 4123, 4138, 4142, 4151, 4162, 4167, 4178, 4186, 4197, 4207, 4212, 4216, 4219, 4221, 4233, 4248, 4280, 4288, 4292, 4296, 4299, 4301, 4310, 4322, 4325, 4348, 4351, 4355, 4376, 4381, 4395, 4397, 4428, 4442, 4454, 4478, 4484, 4515, 4518, 4522, 4526, 4530, 4531, 4537, 4539, 4541, 4545, 4561, 4574, 4646, 4647, 4669, 4672, 4673, 4675, 4678, 4689, 4694, 4736, 4742, 4749, 4753, 4778, 4782, 4783, 4786, 4793, 4794, 4812, 4821, 4823, 4832, 4837, 4840, 4843, 4845, 4846, 4861, 4872, 4897, 4903, 4914, 4927, 4933, 4981, 5009, 5034, 5037, 5038, 5043, 5048, 5086, 5095, 5097, 5115, 5118, 5122, 5123, 5126, 5138, 5139, 5169, 5173, 5174, 5179, 5195, 5197, 5213, 5221, 5223, 5229, 5230, 5254, 5265, 5277, 5291, 5311, 5321, 5327, 5339, 5340, 5341, 5367, 5370, 5374, 5375, 5386, 5391, 5392, 5398, 5433, 5437, 5440, 5466, 5467, 5483, 5488, 5495, 5499, 5503, 5509, 5510, 5517, 5522, 5524, 5541, 5547, 5554, 5586, 5603, 5624, 5631, 5639, 5655, 5657, 5662, 5665, 5692, 5697, 5704, 5726, 5727, 5730, 5745, 5775, 5779, 5780, 5800, 5804, 5809, 5814, 5844, 5853, 5865, 5866, 5881, 5889, 5894, 5895, 5903, 5932, 5939, 5940, 5948, 5950, 5961, 5966, 5974, 5985, 5989, 5994, 5995, 6001, 6030, 6034, 6040, 6057, 6065, 6066, 6075, 6094, 6098, 6129, 6137, 6146, 6157, 6184, 6204, 6206, 6212, 6222, 6223, 6225, 6226, 6237, 6252, 6256, 6260, 6263, 6271, 6278, 6308, 6352, 6358, 6362, 6364, 6376, 6377, 6386, 6388, 6399, 6402, 6414, 6431, 6437, 6449, 6471, 6480, 6482, 6502, 6512, 6516, 6525, 6540, 6542, 6561, 6563, 6568, 6569, 6579, 6600, 6601, 6603, 6605, 6608, 6614, 6647, 6658, 6660, 6668, 6670, 6676, 6679, 6683, 6686, 6688, 6711, 6721, 6725, 6729, 6730, 6737, 6741, 6745, 6793, 6805, 6819, 6821, 6822, 6872, 6877, 6895, 6909, 6915, 6925, 6928, 6935, 6964, 6965, 6977, 6982, 6985, 7010, 7015, 7022, 7030, 7045, 7048, 7069, 7082, 7092, 7096, 7110, 7124, 7129, 7135, 7136, 7142, 7144, 7156, 7162, 7183, 7184, 7197, 7200, 7224, 7227, 7237, 7246, 7255, 7266, 7267, 7297, 7302, 7309, 7324, 7330, 7331, 7336, 7338, 7339, 7341, 7348, 7356, 7371, 7398, 7403, 7404, 7429, 7441, 7443, 7445, 7449, 7463, 7467, 7473, 7521, 7523, 7527, 7531, 7575, 7578, 7580, 7587, 7588, 7592, 7595, 7601, 7613, 7614, 7615, 7618, 7620, 7622, 7630, 7638, 7663, 7688, 7703, 7705, 7715, 7722, 7723, 7734, 7742, 7745, 7749, 7777, 7800, 7816, 7825, 7830, 7831, 7836, 7837, 7851, 7866, 7880, 7886, 7900, 7904, 7926, 7929, 7959, 7965, 7978, 7988, 8000, 8011, 8018, 8030, 8054, 8061, 8062, 8063, 8065, 8072, 8076, 8127, 8147, 8161, 8164, 8193, 8208, 8215, 8225, 8226, 8241, 8260, 8270, 8274, 8278, 8279, 8292, 8304, 8325, 8351, 8372, 8385, 8387, 8396, 8406, 8407, 8410, 8424, 8430, 8446, 8451, 8465, 8470, 8472, 8477, 8486, 8531, 8537, 8538, 8546, 8550, 8564, 8596, 8602, 8606, 8607, 8637, 8643, 8658, 8694, 8697, 8698, 8702, 8711, 8716, 8720, 8736, 8741, 8747, 8757, 8760, 8776, 8779, 8785, 8786, 8789, 8793, 8813, 8820, 8825, 8830, 8853, 8878, 8881, 8898, 8903, 8914, 8915, 8933, 8937, 8941, 8946, 8964, 8974, 8981, 8985, 8998, 9008, 9012, 9019, 9032, 9036, 9044, 9046, 9076, 9101, 9102, 9108, 9112, 9116, 9128, 9144, 9153, 9154, 9155, 9158, 9160, 9174, 9180, 9190, 9194, 9203, 9225, 9231, 9258, 9263, 9271, 9278, 9286, 9290, 9291, 9294, 9313, 9317, 9327, 9328, 9336, 9341, 9361, 9369, 9399, 9408, 9419, 9420, 9421, 9459, 9463, 9467, 9480, 9482, 9483, 9494, 9510, 9539, 9543, 9546, 9551, 9561, 9563, 9570, 9577, 9586, 9590, 9597, 9607, 9615, 9617, 9620, 9628, 9640, 9650, 9659, 9669, 9684, 9685, 9689, 9703, 9719, 9729, 9732, 9735, 9743, 9744, 9751, 9752, 9772, 9779, 9783, 9785, 9786, 9792, 9820, 9821, 9826, 9827, 9829, 9841, 9843, 9853, 9870, 9873, 9881, 9890, 9894, 9895, 9914, 9922, 9947, 9949, 9958, 9972, 9976, 9981, 10001, 10007, 10013, 10016, 10026, 10051, 10061, 10066, 10067, 10075, 10095, 10100, 10114, 10123, 10127, 10149, 10165, 10170, 10174, 10181, 10195, 10201, 10211, 10214, 10220, 10221, 10226, 10229, 10247, 10257, 10262, 10270, 10289, 10303, 10315, 10344, 10345, 10347, 10363, 10369, 10423, 10424, 10425, 10429, 10431, 10434, 10442, 10464, 10469, 10470, 10477, 10480, 10494, 10515, 10516, 10523, 10527, 10528, 10565, 10573, 10578, 10589, 10626, 10638, 10641, 10653, 10655, 10657, 10673, 10674, 10684, 10699, 10706, 10707, 10711, 10731, 10744, 10754, 10763, 10770, 10775, 10782, 10799, 10804, 10805, 10844, 10848, 10851, 10854, 10859, 10862, 10865, 10877, 10878, 10924, 10926, 10928, 10932, 10934, 10939, 10948, 10952, 10953, 10972, 10978, 10986, 10990, 11007, 11015, 11019, 11025, 11033, 11039, 11046, 11050, 11068, 11105, 11126, 11137, 11147, 11149, 11174, 11176, 11179, 11187, 11194, 11214, 11216, 11221, 11232, 11241, 11251, 11252, 11253, 11263, 11291, 11292, 11300, 11311, 11316, 11320, 11330, 11352, 11357, 11358, 11363, 11367, 11376, 11399, 11412, 11413, 11416, 11438, 11441, 11473, 11475, 11490, 11493, 11508, 11556, 11565, 11567, 11568, 11578, 11579, 11580, 11584, 11594, 11606, 11610, 11614, 11617, 11618, 11619, 11625, 11634, 11676, 11700, 11709, 11720, 11732, 11739, 11747, 11755, 11763, 11765, 11767, 11772, 11783, 11790, 11805, 11810, 11823, 11825, 11828, 11830, 11855, 11856, 11887, 11895, 11913, 11931, 11934, 11945, 11973, 11975, 11978, 11986, 11991, 12006, 12018, 12031, 12039, 12049, 12056, 12063, 12064, 12066, 12084, 12086, 12087, 12089, 12100, 12109, 12111, 12120, 12134, 12188, 12189, 12208, 12233, 12268, 12273, 12281, 12282, 12287, 12297, 12307, 12323, 12326, 12336, 12354, 12368, 12396, 12398, 12405, 12424, 12428, 12435, 12462, 12463, 12477, 12487, 12488, 12492, 12523, 12526, 12534, 12548, 12566, 12578, 12587, 12596, 12602, 12607, 12621, 12641, 12646, 12654, 12671, 12682, 12694, 12696, 12704, 12715, 12720, 12724, 12735, 12740, 12754, 12788, 12791, 12807, 12809, 12842, 12853, 12870, 12878, 12886, 12888, 12901, 12914, 12922, 12930, 12944, 12946, 12959, 12977, 12980, 12988, 12994, 12997, 12998, 13002, 13007, 13017, 13049, 13058, 13064, 13087, 13093, 13098, 13101, 13121, 13123, 13132, 13153, 13169, 13170, 13191, 13231, 13234, 13249, 13250, 13252, 13254, 13276, 13295, 13301, 13310, 13311, 13318, 13341, 13342, 13345, 13352, 13370, 13384, 13420, 13429, 13448, 13455, 13462, 13472, 13489, 13504, 13507, 13512, 13515, 13524, 13542, 13546, 13566, 13568, 13577, 13584, 13622, 13652, 13654, 13665, 13671, 13679, 13680, 13696, 13701, 13707, 13716, 13727, 13736, 13744, 13757, 13775, 13790, 13802, 13803, 13815, 13821, 13834, 13837, 13854, 13856, 13861, 13863, 13867, 13869, 13872, 13880, 13884, 13910, 13917, 13919, 13929, 13940, 13941, 13962, 13978, 13981, 13983, 13996, 13998, 14000, 14003, 14015, 14018, 14022, 14027, 14029, 14046, 14081, 14083, 14087, 14088, 14097, 14140, 14146, 14148, 14163, 14179, 14204, 14205, 14244, 14245, 14267, 14273, 14279, 14289, 14303, 14310, 14317, 14318, 14340, 14350, 14360, 14363, 14369, 14376, 14399, 14411, 14432, 14436, 14438, 14454, 14463, 14465, 14487, 14488, 14492, 14503, 14526, 14528, 14536, 14551, 14576, 14608, 14612, 14615, 14620, 14630, 14631, 14638, 14651, 14671, 14678, 14696, 14724, 14727, 14733, 14746, 14754, 14756, 14772, 14774, 14783, 14789, 14791, 14794, 14817, 14843, 14851, 14870, 14887, 14914, 14918, 14921, 14940, 14944, 14947, 14951, 14961, 14967, 14974, 14979, 14982, 14999]

In [0]:

```
#Creating a dataframe with only false positive points
print(type(x_test))

x_test_df = pd.DataFrame(x_test.todense())
print(x_test_df.shape)

x_test_df = x_test_df.iloc[index]
print(x_test_df.shape)
```

```
<class 'scipy.sparse.coo.coo_matrix'>
(15000, 12176)
(1522, 12176)
```

In [0]:

```
#Doing a sum by columns, so that it would be easier to identify the features with zero weightage
x_test_df = x_test_df.sum(axis=0, skipna = True)
print(x_test_df.shape, type(x_test_df))

x_test_df = x_test_df.values.reshape(-1,1)
print(x_test_df.shape, type(x_test_df),x_test_df)
```

```
(12176,) <class 'pandas.core.series.Series'>
(12176, 1) <class 'numpy.ndarray'> [[ 5.]
 [ 5.]
 [80.]
 ...
 [10.]
 [ 2.]
 [ 1.]]
```

In [0]:

```
#Storing all the features with more than 250 weightage in a list

fp_features = []

print(x_test_df.shape[0])
for i in range(x_test_df.shape[0]):
    if x_test_df[i] > 250:
        fp_features.append(feature_names_bow[i])

print(len(fp_features))
```

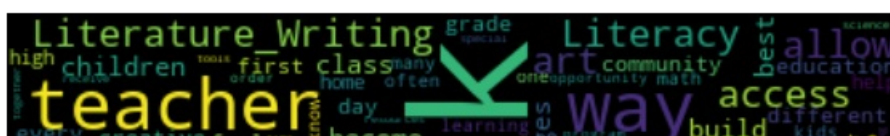
```
12176
157
```

In [0]:

```
#https://www.geeksforgeeks.org/generating-word-cloud-python/
#https://www.programiz.com/python-programming/methods/string/join
#https://www.datacamp.com/community/tutorials/wordcloud-python

from wordcloud import WordCloud

final_fp_features = ",".join(fp_features)
wordcloud = WordCloud(background_color="black").generate(final_fp_features)
plt.figure(figsize=(10,7))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```





In [0]:

```
print(len(index))

print(project_data.columns)
```

1522

```
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects', 'project_is_approved',
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc',
      'senti_score'],
      dtype='object')
```

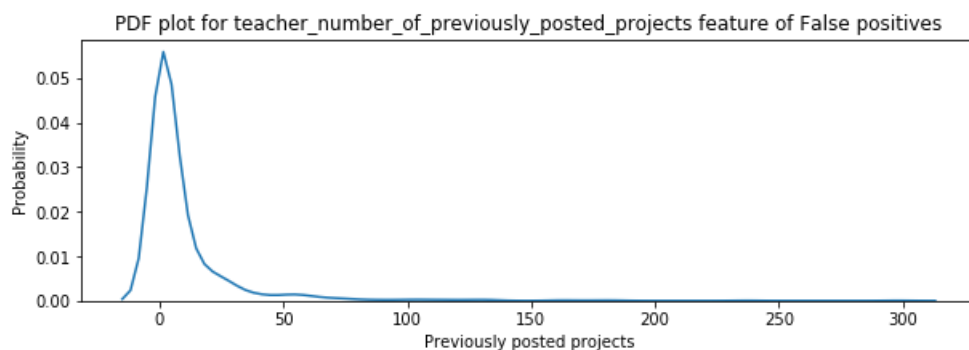
In [0]:

```
#https://www.datacamp.com/community/tutorials/probability-distributions-python
```

```
PDF_df = project_data['teacher_number_of_previously_posted_projects'].iloc[index]
print(type(PDF_df), PDF_df.shape)
```

```
plt.figure(figsize=(10,3))
sns.distplot(PDF_df.values, hist=False)
plt.title('PDF plot for teacher_number_of_previously_posted_projects feature of False positives')
plt.xlabel('Previously posted projects')
plt.ylabel('Probability')
plt.show()
```

<class 'pandas.core.series.Series'> (1522,)



- Most of the projects which were falsely identified as positive, were posted by teachers with very less or zero number of previously posted projects.

## 2.4.2 Applying SGD Classifier brute force on TFIDF, SET 2 (GridSearch)

### Hyper paramter tuning method: GridSearch

In [0]:

```
#https://www.digitalocean.com/community/tutorials/how-to-plot-data-in-python-3-using-matplotlib
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html
#https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter
```

```
from scipy.sparse import hstack
from sklearn.model_selection import GridSearchCV
from sklearn import linear_model
import matplotlib.patches as mpatches
from sklearn.metrics import roc_auc_score
```

```
x_train_tfidf = hstack((categories_one_hot_train, sub_categories_one_hot_train,
grade_cat_one_hot_train,
                        teacher_prefix_one_hot_train, school_state_one_hot_train,
```



```

teacher_prefix_one_hot_train, sub_categories_one_hot_train,
price_train_standardized,
prev_proj_train_standardized, wc_title_train_standardized,
wc_essay_train_standardized, senti_score_train_standardized,
qty_train_standardized, text_train_tfidf, title_train_tfidf))
y_train_tfidf = df_train['project_is_approved']

x_test_tfidf = hstack((categories_one_hot_test, sub_categories_one_hot_test,
grade_cat_one_hot_test,
teacher_prefix_one_hot_test, school_state_one_hot_test, price_test_standardized,
prev_proj_test_standardized, wc_title_test_standardized,
wc_essay_test_standardized, senti_score_test_standardized,
qty_test_standardized, text_test_tfidf, title_test_tfidf))
y_test_tfidf = df_test['project_is_approved']

print(x_train_tfidf.shape, type(x_train_tfidf), y_train_tfidf.shape, type(y_train_tfidf))
print(x_test_tfidf.shape, type(x_test_tfidf), y_test_tfidf.shape, type(y_test_tfidf))

(35000, 12176) <class 'scipy.sparse.coo.coo_matrix'> (35000,) <class 'pandas.core.series.Series'>
(15000, 12176) <class 'scipy.sparse.coo.coo_matrix'> (15000,) <class 'pandas.core.series.Series'>

```

In [0]:

```

#https://stackabuse.com/cross-validation-and-grid-search-for-model-selection-in-python/
#https://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.GridSearchCV.html

#Initialising Classifier
classifier = DecisionTreeClassifier(class_weight='balanced')

#Brute force approach for finding best K value
parameters = {'max_depth': [1, 5, 10, 50, 100, 500, 100],
              'min_samples_split': [5, 10, 100, 500]}

#Training the model on train data
DT_TFIDF = GridSearchCV(classifier, parameters, cv=10, return_train_score=True, scoring='roc_auc',
n_jobs=-1)
DT_TFIDF.fit(x_train_tfidf, y_train_tfidf)

```

Out[0]:

```

GridSearchCV(cv=10, error_score='raise-deprecating',
            estimator=DecisionTreeClassifier(class_weight='balanced',
            criterion='gini', max_depth=None,
            max_features=None,
            max_leaf_nodes=None,
            min_impurity_decrease=0.0,
            min_impurity_split=None,
            min_samples_leaf=1,
            min_samples_split=2,
            min_weight_fraction_leaf=0.0,
            presort=False, random_state=None,
            splitter='best'),
            iid='warn', n_jobs=-1,
            param_grid={'max_depth': [1, 5, 10, 50, 100, 500, 100],
            'min_samples_split': [5, 10, 100, 500]},
            pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
            scoring='roc_auc', verbose=0)

```

In [0]:

```

#https://matplotlib.org/api/\_as\_gen/matplotlib.pyplot.plot.html
#https://stackoverflow.com/questions/20944483/python-3-sort-a-dict-by-its-values/20948781

print(DT_TFIDF.best_params_) #Gives the best value of parameters from the given range
print(DT_TFIDF.cv_results_['params'])

params_train = {}
params_test = {}

for i,j in zip(DT_TFIDF.cv_results_['params'],DT_TFIDF.cv_results_['mean_train_score']):
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))
    params_train[a]=j

for i,i in zip(DT_TFIDF.cv_results_['params'],DT_TFIDF.cv_results_['mean test score']):

```

```

a=(str(i['max_depth'])+', '+str(i['min_samples_split']))
params_test[a]=j

params_train = {k: params_train[k] for k in sorted(params_train)}
params_test = {k: params_test[k] for k in sorted(params_test)}

print(params_train)
print(params_test)

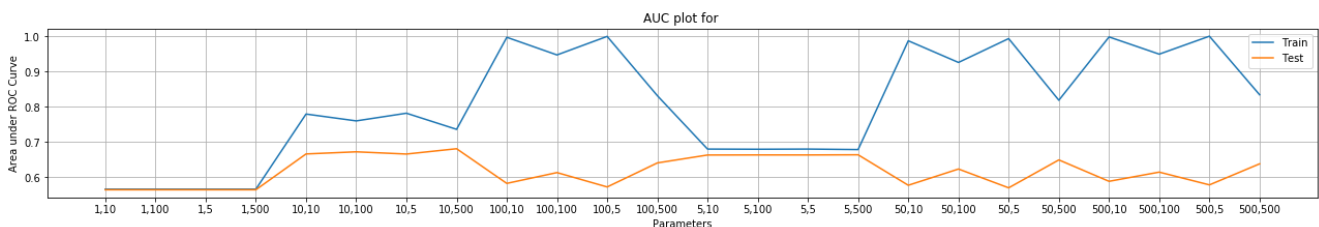
plt.figure(figsize=(22,3))
plt.plot(params_train.keys(),params_train.values(), label="Train")
plt.plot(params_test.keys(),params_test.values(), label="Test")
plt.title('AUC plot for ')
plt.xlabel('Parameters')
plt.ylabel('Area under ROC Curve')
plt.legend()
plt.grid()
plt.show()
plt.close()

```

```

{'max_depth': 10, 'min_samples_split': 500}
[{'max_depth': 1, 'min_samples_split': 5}, {'max_depth': 1, 'min_samples_split': 10},
{'max_depth': 1, 'min_samples_split': 100}, {'max_depth': 1, 'min_samples_split': 500},
{'max_depth': 5, 'min_samples_split': 5}, {'max_depth': 5, 'min_samples_split': 10}, {'max_depth':
5, 'min_samples_split': 100}, {'max_depth': 5, 'min_samples_split': 500}, {'max_depth': 10,
'min_samples_split': 5}, {'max_depth': 10, 'min_samples_split': 10}, {'max_depth': 10,
'min_samples_split': 100}, {'max_depth': 10, 'min_samples_split': 500}, {'max_depth': 50,
'min_samples_split': 5}, {'max_depth': 50, 'min_samples_split': 10}, {'max_depth': 50,
'min_samples_split': 100}, {'max_depth': 50, 'min_samples_split': 500}, {'max_depth': 100,
'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100,
'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}, {'max_depth': 500,
'min_samples_split': 5}, {'max_depth': 500, 'min_samples_split': 10}, {'max_depth': 500,
'min_samples_split': 100}, {'max_depth': 500, 'min_samples_split': 500}, {'max_depth': 100,
'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100,
'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}]
{'1,10': 0.5658447630496513, '1,100': 0.5658447630496513, '1,5': 0.5658447630496513, '1,500': 0.56
58447630496513, '10,10': 0.7790041149886924, '10,100': 0.7595613115623332, '10,5':
0.7813826150347807, '10,500': 0.7355770301342763, '100,10': 0.9968314300989171, '100,100':
0.946771677437496, '100,5': 0.9993509109059684, '100,500': 0.8311269209917208, '5,10':
0.679598389536984, '5,100': 0.6791455017824047, '5,5': 0.6796173471044051, '5,500':
0.6781862048164735, '50,10': 0.9869151939333026, '50,100': 0.9254593650937502, '50,5':
0.9931276554327202, '50,500': 0.8181435325685966, '500,10': 0.9978182294009603, '500,100':
0.9489810922741944, '500,5': 0.999905702924585, '500,500': 0.833743704156233}
{'1,10': 0.5641422578751661, '1,100': 0.5641422578751661, '1,5': 0.5641422578751661, '1,500': 0.56
41422578751661, '10,10': 0.6657641903224826, '10,100': 0.6718169254258323, '10,5':
0.6654909854554135, '10,500': 0.6805807622379173, '100,10': 0.5821878962403411, '100,100':
0.6128886302413883, '100,5': 0.5720998585817175, '100,500': 0.6403000911203013, '5,10':
0.6627287854950775, '5,100': 0.6629807427116557, '5,5': 0.6628963663197197, '5,500':
0.6633888496465025, '50,10': 0.5768960913443291, '50,100': 0.6229874426809158, '50,5':
0.5695971270088239, '50,500': 0.6488472378557606, '500,10': 0.5880832937497642, '500,100':
0.6142040545290168, '500,5': 0.5779854299978125, '500,500': 0.637688299731681}

```



In [0]:

```

#https://datascience.stackexchange.com/questions/22762/understanding-predict-proba-from-
multioutputclassifier
#https://stackoverflow.com/questions/34894587/should-we-plot-the-roc-curve-for-each-class

from sklearn.metrics import roc_curve, auc

#training the model on the best K value found in the above result
final_DT_tfidf = DecisionTreeClassifier(max_depth=10, min_samples_split=500,
class_weight='balanced')
final_DT_tfidf.fit(x_train_tfidf,y_train_tfidf)

```

```

x_train_tfidf_csr=x_train_tfidf.tocsr()
x_test_tfidf_csr=x_test_tfidf.tocsr()

y_train_tfidf_pred=[]
y_test_tfidf_pred=[]

#ROC curve function takes the actual values and the predicted probabilities of the positive class
for i in range(0,x_train_tfidf.shape[0]):
    y_train_tfidf_pred.extend(final_DT_tfidf.predict_proba(x_train_tfidf_csr[i])[:,1]) #[:,1] gives
the probability for class 1

for i in range(0,x_test_tfidf.shape[0]):
    y_test_tfidf_pred.extend(final_DT_tfidf.predict_proba(x_test_tfidf_csr[i])[:,1])

```

In [0]:

```

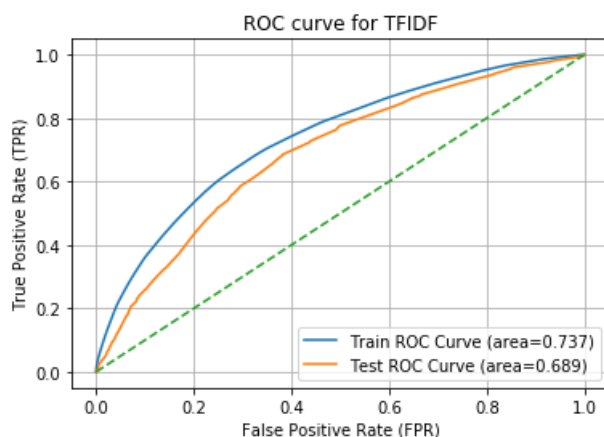
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#Calculating FPR and TPR for train and test data
train_tfidf_fpr, train_tfidf_tpr, train_tfidf_thresholds = roc_curve(y_train_tfidf,
y_train_tfidf_pred)
test_tfidf_fpr, test_tfidf_tpr, test_tfidf_thresholds = roc_curve(y_test_tfidf, y_test_tfidf_pred)

#Calculating AUC for train and test curves
roc_auc_tfidf_train=auc(train_tfidf_fpr,train_tfidf_tpr)
roc_auc_tfidf_test=auc(test_tfidf_fpr,test_tfidf_tpr)

plt.plot(train_tfidf_fpr, train_tfidf_tpr, label="Train ROC Curve (area=%0.3f)" %
roc_auc_tfidf_train)
plt.plot(test_tfidf_fpr, test_tfidf_tpr, label="Test ROC Curve (area=%0.3f)" % roc_auc_tfidf_test)
plt.plot([0,1],[0,1],linestyle='--')
plt.legend()
plt.xlabel("False Positive Rate (FPR)")
plt.ylabel("True Positive Rate (TPR)")
plt.title("ROC curve for TFIDF")
plt.grid()
plt.show()
plt.close()

```



In [0]:

```

#https://stackoverflow.com/questions/35572000/how-can-i-plot-a-confusion-matrix
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
#https://datatofish.com/confusion-matrix-python/

from sklearn.metrics import confusion_matrix as cf_mx

expected_train_tfidf = y_train_tfidf.values
predicted_train_tfidf = final_DT_tfidf.predict(x_train_tfidf)

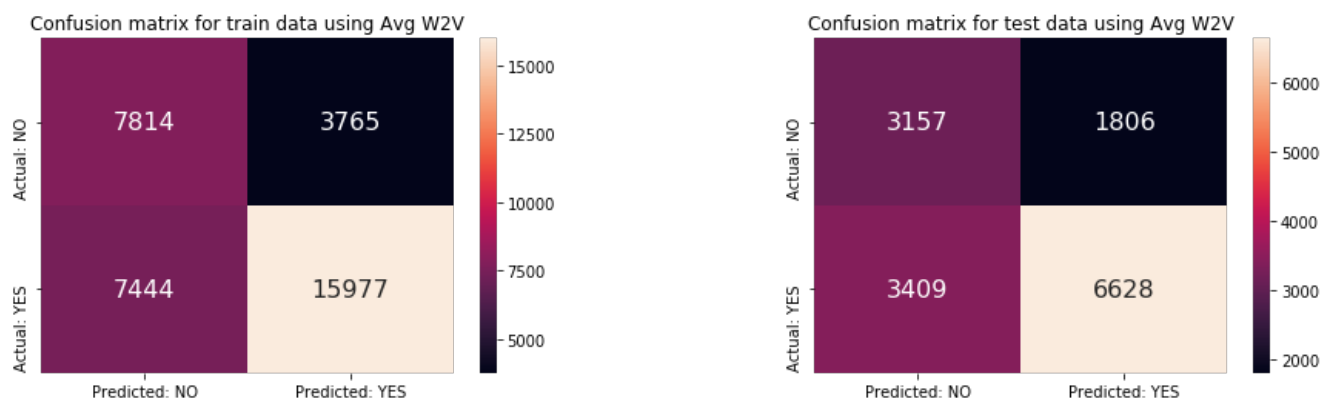
expected_test_tfidf = y_test_tfidf.values
predicted_test_tfidf = final_DT_tfidf.predict(x_test_tfidf)

plt.subplots(figsize=(15,4))

```

```
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_train_tfidf, predicted_train_tfidf)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for train data using Avg W2V')

plt.subplot(1,2,2)
cmdf_test=cf_mx(expected_test_tfidf, predicted_test_tfidf)
df_cm_test = pd.DataFrame(cmdf_test, range(2),range(2))
df_cm_test.columns = ['Predicted: NO','Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for test data using Avg W2V')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()
```



## Decision Tree visualisation using Graphviz

In [0]:

```
#https://scikit-learn.org/stable/modules/generated/sklearn.tree.export_graphviz.html

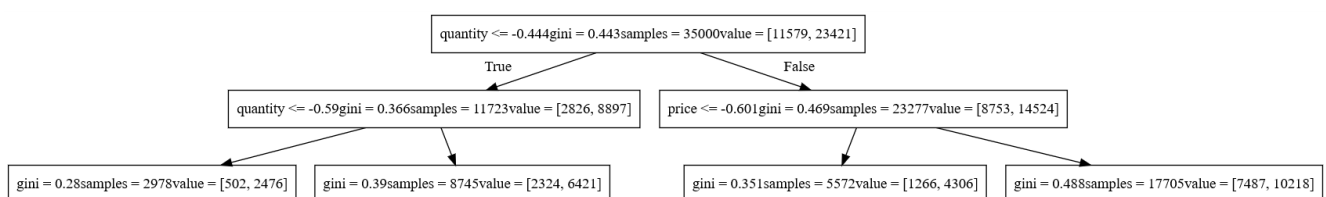
from sklearn import tree

clf = tree.DecisionTreeClassifier(max_depth=2)

clf = clf.fit(x_train_tfidf, y_train_tfidf)
tree.export_graphviz(clf, feature_names=feature_names_tfidf)
```

Out[0]:

```
'digraph Tree {\nnode [shape=box] ;\n0 [label="quantity <= -0.444\\ngini = 0.443\\nsamples = 35000\\nvalue = [11579, 23421]" ] ;\n1 [label="quantity <= -0.59\\ngini = 0.366\\nsamples = 11723\\nvalue = [2826, 8897]" ] ;\n0 -> 1 [labeldistance=2.5, labelangle=45, headlabel="True" ] ;\n2 [label="gini = 0.28\\nsamples = 2978\\nvalue = [502, 2476]" ] ;\n1 -> 2 ;\n3 [label="gini = 0.39\\nsamples = 8745\\nvalue = [2324, 6421]" ] ;\n1 -> 3 ;\n4 [label="price <= -0.601\\ngini = 0.469\\nsamples = 23277\\nvalue = [8753, 14524]" ] ;\n0 -> 4 [labeldistance=2.5, labelangle=-45, headlabel="False" ] ;\n5 [label="gini = 0.351\\nsamples = 5572\\nvalue = [1266, 4306]" ] ;\n4 -> 5 ;\n6 [label="gini = 0.488\\nsamples = 17705\\nvalue = [7487, 10218]" ] ;\n4 -> 6 ;\n}
```



## Word cloud for False Positives

In [0]:

```
#https://stackoverflow.com/questions/36184432/is-it-possible-to-retrieve-false-positives-false-negatives-identified-by-a-conf
#https://stackoverflow.com/questions/31324218/scikit-learn-how-to-obtain-true-positive-true-negative-false-positive-and-fal?rq=1
```

```
print(len(y_test_tfidf),len(predicted_test_tfidf))

index=[]

y_test_tfidf_fp = y_test_tfidf.values.reshape(-1,1)

#Comparing each value of actual y value and predicted y value, in D_Test
#False positive: The actual value is negative(0), but the model predicted it to be positive(1)
for i in range(len(y_test)):
    if (y_test_tfidf_fp[i] == 0) and (predicted_test_tfidf[i] == 1):
        index.append(i) #Storing the index of the FP datapoints in a list

print(len(index))
print(index)
```

15000 15000

1806

[4, 19, 22, 26, 27, 34, 35, 38, 43, 45, 56, 70, 80, 87, 118, 119, 123, 132, 159, 168, 169, 185, 19  
2, 200, 203, 232, 235, 250, 299, 316, 317, 320, 328, 331, 338, 344, 352, 353, 361, 378, 384, 386,  
388, 391, 401, 407, 414, 436, 440, 444, 454, 460, 462, 467, 471, 488, 495, 511, 517, 525, 528, 529  
, 535, 537, 561, 563, 565, 583, 586, 601, 603, 604, 608, 613, 623, 642, 648, 660, 672, 676, 678, 6  
83, 696, 700, 713, 722, 724, 731, 736, 743, 749, 754, 767, 774, 792, 796, 801, 806, 807, 809, 819,  
821, 832, 843, 849, 860, 863, 865, 872, 892, 896, 931, 944, 967, 983, 986, 1003, 1013, 1022, 1048,  
1062, 1064, 1071, 1084, 1103, 1105, 1150, 1151, 1152, 1158, 1206, 1208, 1212, 1220, 1234, 1243, 124  
9, 1250, 1251, 1254, 1258, 1267, 1275, 1277, 1290, 1313, 1319, 1326, 1332, 1340, 1341, 1349, 1358,  
1369, 1375, 1384, 1397, 1423, 1453, 1460, 1493, 1503, 1507, 1514, 1526, 1541, 1544, 1555, 1557, 156  
3, 1574, 1576, 1600, 1606, 1610, 1637, 1640, 1650, 1659, 1664, 1666, 1672, 1674, 1685, 1686, 1693,  
1706, 1743, 1747, 1761, 1767, 1768, 1793, 1797, 1800, 1812, 1820, 1847, 1848, 1849, 1864, 1879, 188  
5, 1887, 1890, 1902, 1906, 1916, 1917, 1920, 1922, 1932, 1935, 1939, 1993, 2009, 2016, 2017, 2026,  
2037, 2059, 2061, 2062, 2083, 2101, 2105, 2107, 2108, 2117, 2123, 2126, 2127, 2128, 2131, 2151, 215  
3, 2157, 2179, 2200, 2207, 2211, 2215, 2216, 2226, 2231, 2244, 2269, 2270, 2272, 2274, 2282, 2296,  
2308, 2317, 2322, 2328, 2358, 2371, 2373, 2380, 2397, 2406, 2422, 2428, 2431, 2432, 2436, 2442, 244  
3, 2448, 2450, 2453, 2476, 2478, 2481, 2488, 2490, 2494, 2505, 2509, 2516, 2518, 2524, 2526, 2530,  
2546, 2554, 2567, 2578, 2580, 2583, 2592, 2615, 2622, 2626, 2632, 2633, 2644, 2653, 2656, 2663, 266  
6, 2670, 2674, 2681, 2685, 2686, 2711, 2712, 2720, 2723, 2735, 2737, 2748, 2754, 2762, 2767, 2780,  
2791, 2793, 2801, 2811, 2813, 2825, 2836, 2839, 2844, 2859, 2885, 2886, 2902, 2927, 2931, 2937, 295  
3, 2955, 2957, 2978, 2980, 2994, 3000, 3003, 3014, 3020, 3027, 3051, 3064, 3081, 3083, 3089, 3097,  
3108, 3114, 3119, 3120, 3126, 3138, 3142, 3149, 3158, 3197, 3202, 3203, 3208, 3211, 3213, 3224, 322  
8, 3230, 3231, 3239, 3248, 3250, 3283, 3295, 3304, 3308, 3314, 3327, 3334, 3339, 3344, 3350, 3367,  
3386, 3427, 3442, 3444, 3446, 3455, 3457, 3464, 3471, 3475, 3479, 3483, 3507, 3513, 3515, 3524, 352  
6, 3527, 3538, 3545, 3550, 3555, 3559, 3561, 3580, 3581, 3597, 3598, 3603, 3605, 3607, 3608, 3614,  
3616, 3617, 3628, 3640, 3645, 3657, 3672, 3682, 3687, 3723, 3733, 3734, 3741, 3742, 3748, 3753, 375  
8, 3768, 3770, 3774, 3775, 3789, 3793, 3805, 3807, 3818, 3819, 3821, 3822, 3824, 3829, 3832, 3834,  
3839, 3844, 3861, 3864, 3865, 3904, 3910, 3926, 3936, 3941, 3945, 3946, 3952, 3961, 3970, 3972, 397  
3, 3975, 3981, 3994, 3996, 4002, 4003, 4004, 4019, 4023, 4028, 4030, 4033, 4076, 4084, 4100, 4112,  
4123, 4138, 4142, 4151, 4162, 4165, 4178, 4181, 4185, 4197, 4204, 4207, 4210, 4212, 4216, 4219, 422  
1, 4233, 4241, 4248, 4252, 4280, 4288, 4292, 4296, 4301, 4306, 4310, 4322, 4325, 4348, 4351, 4355,  
4370, 4376, 4381, 4395, 4397, 4406, 4428, 4442, 4454, 4478, 4482, 4484, 4503, 4504, 4510, 4515, 452  
2, 4526, 4531, 4537, 4539, 4541, 4545, 4554, 4561, 4568, 4571, 4574, 4594, 4646, 4669, 4673, 4675,  
4676, 4678, 4689, 4694, 4700, 4726, 4736, 4742, 4749, 4753, 4768, 4778, 4782, 4783, 4786, 4793, 479  
4, 4807, 4810, 4823, 4832, 4837, 4840, 4843, 4845, 4846, 4856, 4861, 4872, 4897, 4903, 4931, 4933,  
4941, 4946, 4952, 4981, 4993, 5009, 5018, 5034, 5037, 5038, 5043, 5048, 5070, 5086, 5095, 5115, 511  
8, 5122, 5123, 5126, 5130, 5131, 5138, 5139, 5169, 5170, 5173, 5174, 5178, 5179, 5184, 5193, 5195,  
5197, 5213, 5220, 5221, 5223, 5229, 5230, 5247, 5254, 5258, 5265, 5277, 5280, 5291, 5311, 5320, 532  
1, 5339, 5340, 5341, 5367, 5370, 5374, 5375, 5386, 5391, 5392, 5398, 5420, 5433, 5437, 5440, 5454,  
5466, 5467, 5483, 5488, 5492, 5501, 5503, 5510, 5517, 5524, 5541, 5547, 5554, 5582, 5586, 5607, 562  
4, 5631, 5639, 5655, 5657, 5662, 5665, 5692, 5694, 5697, 5704, 5726, 5727, 5730, 5741, 5745, 5762,  
5769, 5779, 5780, 5800, 5802, 5809, 5814, 5823, 5844, 5851, 5853, 5865, 5866, 5867, 5873, 5881, 588  
9, 5895, 5900, 5903, 5915, 5918, 5932, 5939, 5940, 5948, 5950, 5961, 5966, 5974, 5985, 5989, 5994,  
5998, 6001, 6006, 6024, 6030, 6034, 6040, 6057, 6065, 6066, 6072, 6075, 6077, 6086, 6090, 6094, 609  
8, 6129, 6137, 6143, 6146, 6177, 6184, 6195, 6204, 6206, 6212, 6221, 6222, 6223, 6225, 6252, 6256,  
6258, 6260, 6263, 6271, 6272, 6281, 6303, 6308, 6317, 6344, 6352, 6357, 6360, 6362, 6364, 6376, 637  
7, 6388, 6390, 6399, 6402, 6404, 6414, 6418, 6421, 6429, 6431, 6437, 6449, 6480, 6482, 6498, 6505,  
6512, 6516, 6525, 6540, 6542, 6557, 6561, 6563, 6564, 6568, 6569, 6579, 6581, 6600, 6601, 6603, 660  
5, 6608, 6614, 6631, 6632, 6647, 6658, 6668, 6670, 6676, 6679, 6683, 6688, 6711, 6721, 6725, 6729,  
6730, 6741, 6745, 6746, 6754, 6794, 6805, 6808, 6819, 6821, 6822, 6834, 6872, 6877, 6882, 6890, 689  
3, 6895, 6909, 6915, 6924, 6925, 6935, 6964, 6965, 6977, 6982, 6985, 6988, 6998, 7003, 7015, 7022,  
7036, 7045, 7048, 7049, 7056, 7069, 7082, 7092, 7096, 7110, 7123, 7129, 7133, 7135, 7136, 7142, 714  
4, 7183, 7197, 7200, 7224, 7227, 7232, 7237, 7240, 7246, 7255, 7266, 7267, 7272, 7274, 7290, 7291,  
7302, 7324, 7330, 7331, 7336, 7338, 7341, 7348, 7356, 7377, 7395, 7398, 7403, 7404, 7429, 7432, 744  
6, 7456, 7467, 7470, 7472, 7511, 7521, 7522, 7523, 7527, 7531, 7544, 7575, 7576, 7580, 7582, 7583, 7588

9, 1459, 1461, 1413, 1511, 1521, 1523, 1521, 1531, 1544, 1515, 1518, 1580, 1581, 1588, 1592, 1599, 7601, 7613, 7615, 7618, 7620, 7621, 7622, 7632, 7638, 7655, 7660, 7663, 7683, 7703, 7705, 7715, 7718, 7721, 7722, 7723, 7734, 7742, 7749, 7750, 7777, 7788, 7790, 7800, 7816, 7825, 7830, 7831, 7836, 7837, 7851, 7866, 7880, 7886, 7900, 7904, 7924, 7926, 7959, 7964, 7965, 7967, 7978, 7982, 7988, 8000, 8011, 8018, 8020, 8030, 8054, 8055, 8061, 8062, 8063, 8065, 8072, 8076, 8085, 8127, 8147, 8153, 8161, 8164, 8176, 8193, 8205, 8208, 8210, 8211, 8214, 8215, 8225, 8240, 8241, 8260, 8270, 8274, 8278, 8279, 8292, 8325, 8351, 8358, 8372, 8377, 8385, 8387, 8390, 8396, 8406, 8407, 8410, 8424, 8430, 8446, 8451, 8465, 8467, 8470, 8472, 8473, 8477, 8486, 8531, 8537, 8538, 8544, 8546, 8550, 8552, 8568, 8602, 8606, 8607, 8637, 8653, 8682, 8694, 8697, 8698, 8702, 8710, 8719, 8720, 8736, 8741, 8747, 8754, 8757, 8760, 8767, 8776, 8779, 8785, 8786, 8789, 8792, 8793, 8813, 8825, 8830, 8853, 8878, 8880, 8881, 8903, 8915, 8933, 8936, 8937, 8941, 8946, 8964, 8974, 8981, 8998, 9008, 9016, 9019, 9032, 9044, 9046, 9086, 9102, 9107, 9108, 9112, 9116, 9128, 9140, 9153, 9154, 9155, 9158, 9160, 9174, 9175, 9180, 9183, 9190, 9194, 9225, 9238, 9240, 9245, 9256, 9258, 9263, 9271, 9278, 9286, 9290, 9294, 9304, 9313, 9317, 9327, 9328, 9336, 9341, 9354, 9369, 9375, 9399, 9407, 9408, 9418, 9419, 9420, 9421, 9423, 9425, 9431, 9463, 9467, 9480, 9482, 9494, 9508, 9510, 9539, 9543, 9546, 9551, 9552, 9557, 9563, 9570, 9574, 9577, 9590, 9597, 9607, 9614, 9615, 9620, 9628, 9640, 9659, 9661, 9669, 9684, 9685, 9689, 9703, 9711, 9719, 9721, 9729, 9735, 9743, 9744, 9751, 9752, 9779, 9783, 9785, 9808, 9820, 9821, 9823, 9826, 9827, 9829, 9841, 9851, 9853, 9856, 9870, 9880, 9881, 9882, 9890, 9894, 9895, 9900, 9903, 9922, 9947, 9958, 9969, 9976, 9977, 9981, 9993, 10001, 10013, 10016, 10025, 10026, 10051, 10066, 10067, 10073, 10075, 10090, 10100, 10114, 10127, 10149, 10164, 10165, 10168, 10170, 10174, 10179, 10181, 10192, 10201, 10211, 10214, 10220, 10221, 10226, 10229, 10235, 10247, 10257, 10260, 10262, 10270, 10280, 10289, 10299, 10303, 10315, 10344, 10345, 10347, 10357, 10358, 10363, 10369, 10376, 10378, 10390, 10401, 10423, 10424, 10429, 10431, 10434, 10447, 10449, 10464, 10480, 10494, 10511, 10515, 10523, 10527, 10528, 10561, 10573, 10578, 10589, 10612, 10625, 10626, 10638, 10641, 10651, 10653, 10659, 10673, 10674, 10684, 10697, 10706, 10707, 10711, 10731, 10753, 10754, 10758, 10763, 10770, 10775, 10776, 10782, 10799, 10800, 10804, 10805, 10812, 10820, 10844, 10848, 10851, 10854, 10859, 10862, 10865, 10877, 10878, 10882, 10899, 10915, 10918, 10924, 10932, 10934, 10939, 10945, 10948, 10952, 10958, 10962, 10972, 10973, 10986, 10990, 11007, 11015, 11019, 11025, 11033, 11046, 11050, 11110, 11124, 11125, 11126, 11130, 11137, 11149, 11174, 11176, 11179, 11187, 11194, 11204, 11216, 11217, 11221, 11232, 11241, 11251, 11252, 11253, 11263, 11291, 11292, 11294, 11305, 11311, 11316, 11320, 11330, 11352, 11357, 11358, 11363, 11367, 11376, 11392, 11399, 11412, 11413, 11418, 11438, 11441, 11457, 11473, 11475, 11482, 11490, 11491, 11493, 11508, 11531, 11556, 11565, 11567, 11568, 11578, 11579, 11584, 11593, 11594, 11605, 11617, 11618, 11621, 11625, 11634, 11657, 11668, 11676, 11678, 11680, 11691, 11700, 11709, 11715, 11720, 11732, 11739, 11747, 11753, 11755, 11765, 11767, 11774, 11776, 11783, 11792, 11805, 11806, 11810, 11811, 11825, 11828, 11830, 11832, 11855, 11856, 11858, 11871, 11887, 11895, 11902, 11913, 11931, 11934, 11945, 11947, 11955, 11970, 11973, 11975, 11978, 11986, 11991, 12004, 12006, 12013, 12022, 12031, 12049, 12060, 12063, 12064, 12066, 12086, 12087, 12089, 12100, 12111, 12120, 12134, 12138, 12148, 12183, 12188, 12189, 12197, 12200, 12208, 12245, 12268, 12273, 12278, 12281, 12282, 12287, 12297, 12307, 12322, 12323, 12326, 12336, 12351, 12354, 12357, 12396, 12398, 12405, 12424, 12428, 12435, 12439, 12462, 12463, 12481, 12487, 12488, 12492, 12496, 12500, 12504, 12507, 12523, 12525, 12526, 12534, 12557, 12566, 12578, 12587, 12590, 12596, 12607, 12610, 12611, 12621, 12625, 12636, 12641, 12669, 12671, 12682, 12692, 12696, 12698, 12704, 12706, 12707, 12715, 12720, 12724, 12731, 12740, 12754, 12788, 12791, 12807, 12809, 12842, 12853, 12870, 12875, 12878, 12886, 12887, 12888, 12914, 12922, 12930, 12944, 12959, 12962, 12967, 12973, 12980, 12988, 12994, 12997, 13002, 13007, 13009, 13017, 13038, 13058, 13064, 13087, 13093, 13098, 13101, 13110, 13121, 13132, 13144, 13153, 13167, 13170, 13175, 13191, 13228, 13234, 13241, 13249, 13250, 13252, 13254, 13273, 13274, 13276, 13279, 13289, 13291, 13292, 13295, 13301, 13310, 13318, 13341, 13342, 13345, 13347, 13352, 13370, 13384, 13426, 13429, 13435, 13442, 13448, 13455, 13462, 13472, 13479, 13489, 13504, 13507, 13524, 13525, 13542, 13543, 13546, 13566, 13577, 13580, 13584, 13588, 13617, 13622, 13651, 13652, 13654, 13658, 13665, 13671, 13673, 13679, 13680, 13685, 13696, 13697, 13701, 13706, 13716, 13727, 13736, 13744, 13748, 13757, 13775, 13782, 13790, 13802, 13803, 13815, 13826, 13832, 13834, 13837, 13854, 13856, 13861, 13863, 13867, 13869, 13872, 13880, 13884, 13919, 13929, 13940, 13950, 13962, 13964, 13969, 13981, 13983, 13996, 13998, 14000, 14003, 14015, 14018, 14022, 14027, 14029, 14043, 14046, 14059, 14081, 14083, 14084, 14087, 14088, 14096, 14097, 14109, 14136, 14140, 14146, 14148, 14163, 14179, 14188, 14204, 14205, 14240, 14244, 14245, 14260, 14267, 14273, 14279, 14289, 14303, 14310, 14317, 14318, 14322, 14340, 14350, 14360, 14369, 14376, 14399, 14411, 14415, 14432, 14436, 14437, 14438, 14447, 14453, 14454, 14463, 14465, 14487, 14488, 14491, 14492, 14503, 14514, 14526, 14528, 14541, 14551, 14576, 14578, 14583, 14585, 14608, 14612, 14615, 14620, 14630, 14631, 14638, 14651, 14653, 14665, 14666, 14678, 14696, 14700, 14707, 14724, 14727, 14733, 14739, 14746, 14753, 14754, 14772, 14774, 14777, 14783, 14784, 14789, 14791, 14794, 14817, 14824, 14827, 14842, 14843, 14851, 14870, 14887, 14914, 14918, 14921, 14925, 14940, 14944, 14947, 14951, 14967, 14973, 14974, 14982, 14999]

In [0]:

```
#Creating a dataframe with only false positive points
print(type(x_test_tfidf))

x_test_tfidf_df = pd.DataFrame(x_test_tfidf.todense())
print(x_test_tfidf_df.shape)

x_test_tfidf_df = x_test_tfidf_df.iloc[index]
print(x_test_tfidf_df.shape)
```

```
<class 'scipy.sparse.coo.coo_matrix'>
(15000, 12176)
```

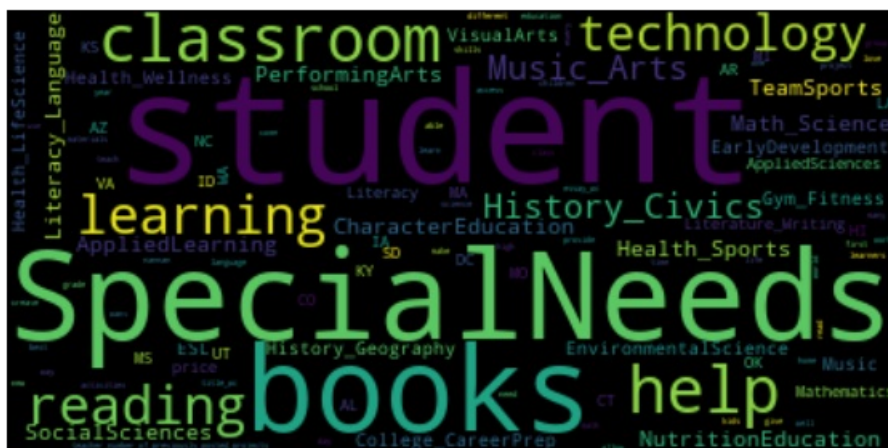
In [0]:

```
x_test_tfidf_df = x_test_tfidf_df.values.reshape(-1,1)
print(x_test_tfidf_df.shape, type(x_test_tfidf_df), x_test_tfidf_df)
```

In [0]:

12176  
124

```
final_fp_features = ",".join(fp_features)
wordcloud = WordCloud(background_color="black").generate(final_fp_features)
plt.figure(figsize=(10,7))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



## Box plot for Price feature of False positives

In [0]:

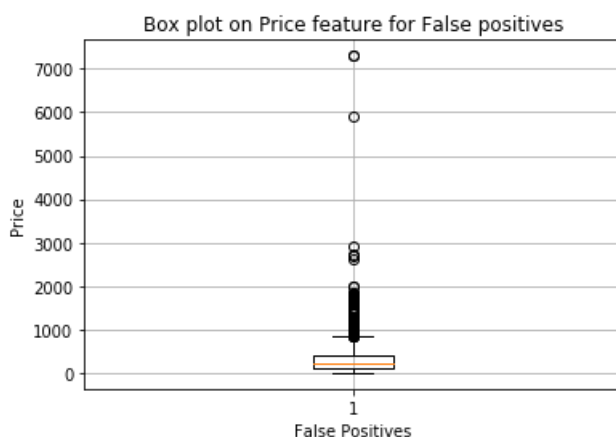
```
print(len(index))  
  
print(project_data.columns)
```

```
1806  
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',  
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',  
      'project_essay_2', 'project_essay_3', 'project_essay_4',  
      'project_resource_summary',  
      'teacher_number_of_previously_posted_projects', 'project_is_approved',  
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',  
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc',  
      'senti_score'],  
      dtype='object')
```

In [0]:

```
BP_df = project_data['price'].iloc[index]  
print(type(BP_df), BP_df.shape)  
  
plt.boxplot(BP_df.values)  
plt.title('Box plot on Price feature for False positives')  
plt.xlabel('False Positives')  
plt.ylabel('Price')  
plt.grid()  
plt.show()
```

<class 'pandas.core.series.Series'> (1806,)



- Most of the projects which have been falsely identified as positive are below 500 dollars.

## PDF plot for 'teacher\_number\_of\_previously\_posted\_projects' feature of False positives

In [0]:

```
print(len(index))  
  
print(project_data.columns)
```

```
1806  
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',  
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',  
      'project_essay_2', 'project_essay_3', 'project_essay_4',  
      'project_resource_summary',  
      'teacher_number_of_previously_posted_projects', 'project_is_approved',  
      'clean_categories', 'clean_subcategories', 'essay', 'clean_essay',  
      'clean_title', 'price', 'quantity', 'title_wc', 'essay_wc',  
      'senti_score']
```



```
senti_score'],
dtype='object')
```

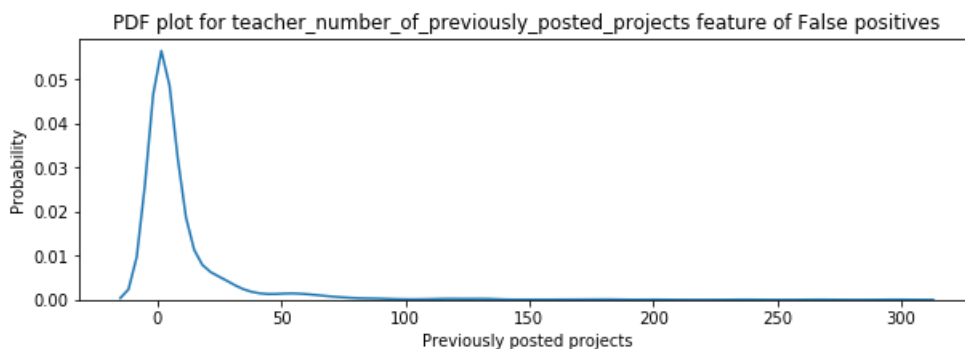
In [0]:

```
#https://www.datacamp.com/community/tutorials/probability-distributions-python
```

```
PDF_df = project_data['teacher_number_of_previously_posted_projects'].iloc[index]
print(type(PDF_df), PDF_df.shape)
```

```
plt.figure(figsize=(10,3))
sns.distplot(PDF_df.values, hist=False)
plt.title('PDF plot for teacher_number_of_previously_posted_projects feature of False positives')
plt.xlabel('Previously posted projects')
plt.ylabel('Probability')
plt.show()
```

```
<class 'pandas.core.series.Series'> (1806,)
```



- Most of the projects which have been falsely identified as positive, were posted by teachers with very less number of previously posted projects.

### 2.4.3 Applying SGD Classifier brute force on AVG W2V, SET 3

#### Hyper paramter tuning method: GridSearch

In [0]:

```
#https://www.digitalocean.com/community/tutorials/how-to-plot-data-in-python-3-using-matplotlib
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html
#https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter
```

```
from scipy.sparse import hstack
from sklearn.model_selection import GridSearchCV
from sklearn import linear_model
import matplotlib.patches as mpatches
from sklearn.metrics import roc_auc_score

x_train_avg_w2v = hstack((categories_one_hot_train, sub_categories_one_hot_train,
grade_cat_one_hot_train,
teacher_prefix_one_hot_train, school_state_one_hot_train,
price_train_standardized,
prev_proj_train_standardized, wc_title_train_standardized,
wc_essay_train_standardized, senti_score_train_standardized,
qty_train_standardized, avg_w2v_train_text_vectors,
avg_w2v_title_train_vectors))
y_train_avg_w2v = df_train['project_is_approved']

x_test_avg_w2v = hstack((categories_one_hot_test, sub_categories_one_hot_test,
grade_cat_one_hot_test,
teacher_prefix_one_hot_test, school_state_one_hot_test, price_test_standardized,
prev_proj_test_standardized, wc_title_test_standardized,
wc_essay_test_standardized, senti_score_test_standardized,
qty_test_standardized, avg_w2v_test_text_vectors,
avg_w2v_title_test_vectors))
y_test_avg_w2v = df_test['project_is_approved']
```

```
print(x_train_avg_w2v.shape, type(x_train_avg_w2v), y_train_avg_w2v.shape, type(y_train_avg_w2v))
print(x_test_avg_w2v.shape, type(x_test_avg_w2v), y_test_avg_w2v.shape, type(y_test_avg_w2v))
```

```
(35000, 206) <class 'scipy.sparse.coo.coo_matrix'> (35000,) <class 'pandas.core.series.Series'>
(15000, 206) <class 'scipy.sparse.coo.coo_matrix'> (15000,) <class 'pandas.core.series.Series'>
```

In [0]:

```
#https://stackabuse.com/cross-validation-and-grid-search-for-model-selection-in-python/
#https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

#Initialising Classifier
classifier = DecisionTreeClassifier(class_weight='balanced')

#Brute force approach for finding best K value
parameters = {'max_depth': [1, 5, 10, 50, 100, 500, 1000],
              'min_samples_split': [5, 10, 100, 500]}

#Training the model on train data
DT_avg_w2v = GridSearchCV(classifier, parameters, return_train_score=True, cv=10,
scoring='roc_auc', n_jobs=-1)
DT_avg_w2v.fit(x_train_avg_w2v, y_train_avg_w2v)
```

Out [0]:

```
GridSearchCV(cv=10, error_score='raise-deprecating',
            estimator=DecisionTreeClassifier(class_weight='balanced',
                                             criterion='gini', max_depth=None,
                                             max_features=None,
                                             max_leaf_nodes=None,
                                             min_impurity_decrease=0.0,
                                             min_impurity_split=None,
                                             min_samples_leaf=1,
                                             min_samples_split=2,
                                             min_weight_fraction_leaf=0.0,
                                             presort=False, random_state=None,
                                             splitter='best'),
            iid='warn', n_jobs=-1,
            param_grid={'max_depth': [1, 5, 10, 50, 100, 500, 1000],
                       'min_samples_split': [5, 10, 100, 500]},
            pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
            scoring='roc_auc', verbose=0)
```

In [0]:

```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://stackoverflow.com/questions/20944483/python-3-sort-a-dict-by-its-values/20948781

print(DT_avg_w2v.best_params_) #Gives the best value of parameters from the given range
print(DT_avg_w2v.cv_results_['params'])

params_train = {}
params_test = {}

for i,j in zip(DT_avg_w2v.cv_results_['params'],DT_avg_w2v.cv_results_['mean_train_score']):
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))
    params_train[a]=j

for i,j in zip(DT_avg_w2v.cv_results_['params'],DT_avg_w2v.cv_results_['mean_test_score']):
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))
    params_test[a]=j

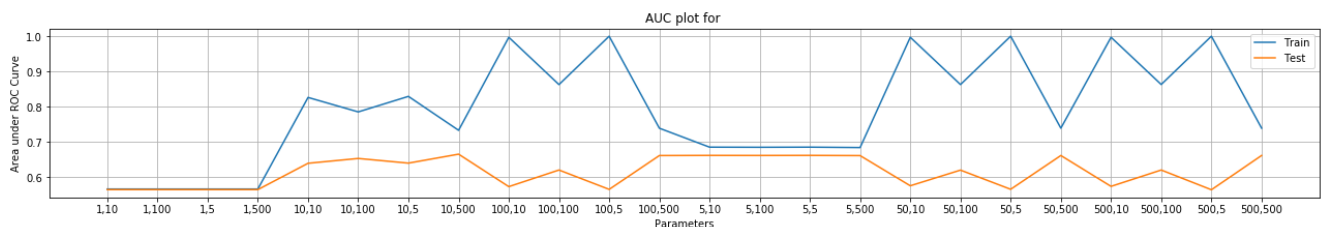
params_train = {k: params_train[k] for k in sorted(params_train)}
params_test = {k: params_test[k] for k in sorted(params_test)}

print(params_train)
print(params_test)

plt.figure(figsize=(22,3))
plt.plot(params_train.keys(),params_train.values(), label="Train")
plt.plot(params_test.keys(),params_test.values(), label="Test")
plt.title('AUC plot for ')
plt.xlabel('Parameters')
```

```
plt.ylabel('Area under ROC Curve')
plt.legend()
plt.grid()
plt.show()
plt.close()
```

```
{'max_depth': 10, 'min_samples_split': 500}
[{'max_depth': 1, 'min_samples_split': 5}, {'max_depth': 1, 'min_samples_split': 10},
{'max_depth': 1, 'min_samples_split': 100}, {'max_depth': 1, 'min_samples_split': 500},
{'max_depth': 5, 'min_samples_split': 5}, {'max_depth': 5, 'min_samples_split': 10}, {'max_depth': 5, 'min_samples_split': 100}, {'max_depth': 5, 'min_samples_split': 500}, {'max_depth': 10, 'min_samples_split': 5}, {'max_depth': 10, 'min_samples_split': 10}, {'max_depth': 10, 'min_samples_split': 100}, {'max_depth': 10, 'min_samples_split': 500}, {'max_depth': 50, 'min_samples_split': 5}, {'max_depth': 50, 'min_samples_split': 10}, {'max_depth': 50, 'min_samples_split': 100}, {'max_depth': 50, 'min_samples_split': 500}, {'max_depth': 100, 'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100, 'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}, {'max_depth': 500, 'min_samples_split': 5}, {'max_depth': 500, 'min_samples_split': 10}, {'max_depth': 500, 'min_samples_split': 100}, {'max_depth': 500, 'min_samples_split': 500}, {'max_depth': 100, 'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100, 'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}]
{'1,10': 0.5658447630496513, '1,100': 0.5658447630496513, '1,5': 0.5658447630496513, '1,500': 0.5658447630496513, '10,10': 0.825948331959301, '10,100': 0.7848489980518301, '10,5': 0.8289124129794793, '10,500': 0.7325695616021869, '100,10': 0.9967880431940497, '100,100': 0.8624575526496674, '100,5': 0.9998384321037822, '100,500': 0.7385302096706652, '5,10': 0.6847673431950698, '5,100': 0.6843965871460929, '5,5': 0.6847673431950698, '5,500': 0.6836220408587543, '50,10': 0.9968111355585707, '50,100': 0.862508372975855, '50,5': 0.9998332717860254, '50,500': 0.7385938128715847, '500,10': 0.9967735067307275, '500,100': 0.8627068395950728, '500,5': 0.9998362449057137, '500,500': 0.7386429385238374}
{'1,10': 0.5641422578751661, '1,100': 0.5641422578751661, '1,5': 0.5641422578751661, '1,500': 0.5641422578751661, '10,10': 0.6389473370892925, '10,100': 0.6526521670726442, '10,5': 0.6393735020875281, '10,500': 0.6648379438349935, '100,10': 0.5726117343619297, '100,100': 0.6198176303633686, '100,5': 0.5648376504922946, '100,500': 0.6609035182506767, '5,10': 0.6613000823125795, '5,100': 0.6611194264671314, '5,5': 0.6613427808608289, '5,500': 0.6607477217161226, '50,10': 0.5750594475062573, '50,100': 0.6195481510640666, '50,5': 0.5650829458199105, '50,500': 0.6610182659900135, '500,10': 0.5733789780843133, '500,100': 0.6198076430866286, '500,5': 0.5637514995833907, '500,500': 0.6611120700545612}
```



In [0]:

```
#https://datascience.stackexchange.com/questions/22762/understanding-predict-proba-from-multioutputclassifier
#https://stackoverflow.com/questions/34894587/should-we-plot-the-roc-curve-for-each-class
#https://stackoverflow.com/questions/55250963/how-to-get-probabilities-for-sgdclassifier-linearsvm

from sklearn.metrics import roc_curve, auc

#training the model on the best K value found in the above result
final_DT_avg_w2v = DecisionTreeClassifier(max_depth=10, min_samples_split=500,
class_weight='balanced')
final_DT_avg_w2v.fit(x_train_avg_w2v, y_train_avg_w2v)

x_train_avg_w2v_csr=x_train_avg_w2v.tocsr()
x_test_avg_w2v_csr=x_test_avg_w2v.tocsr()

y_train_avg_w2v_pred=[]
y_test_avg_w2v_pred=[]

#ROC curve function takes the actual values and the predicted probabilities of the positive class
for i in range(0,x_train_avg_w2v.shape[0]):
    y_train_avg_w2v_pred.extend(final_DT_avg_w2v.predict_proba(x_train_avg_w2v_csr[i])[:,1]) #[:,1]
gives the probability for class 1

for i in range(0,x_test_avg_w2v.shape[0]):
    y_test_avg_w2v_pred.extend(final_DT_avg_w2v.predict_proba(x_test_avg_w2v_csr[i])[:,1])
```

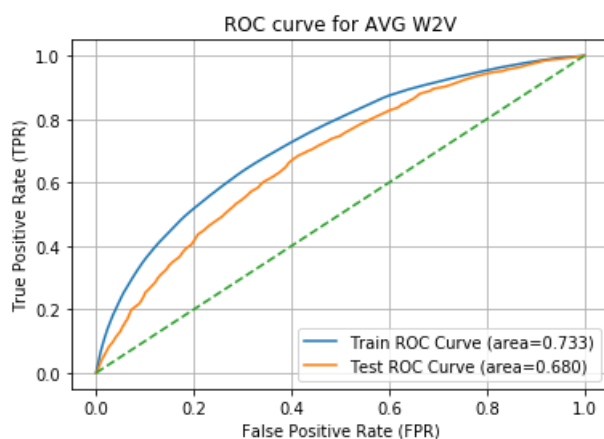
In [0]:

```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#Calculating FPR and TPR for train and test data
train_avg_w2v_fpr, train_avg_w2v_tpr, train_avg_w2v_thresholds = roc_curve(y_train_avg_w2v,
y_train_avg_w2v_pred)
test_avg_w2v_fpr, test_avg_w2v_tpr, test_avg_w2v_thresholds = roc_curve(y_test_avg_w2v, y_test_avg_w2v_pred)

#Calculating AUC for train and test curves
roc_auc_avg_w2v_train=auc(train_avg_w2v_fpr,train_avg_w2v_tpr)
roc_auc_avg_w2v_test=auc(test_avg_w2v_fpr,test_avg_w2v_tpr)

plt.plot(train_avg_w2v_fpr, train_avg_w2v_tpr, label="Train ROC Curve (area=%0.3f)" %
roc_auc_avg_w2v_train)
plt.plot(test_avg_w2v_fpr, test_avg_w2v_tpr, label="Test ROC Curve (area=%0.3f)" %
roc_auc_avg_w2v_test)
plt.plot([0,1],[0,1],linestyle='--')
plt.legend()
plt.xlabel("False Positive Rate (FPR)")
plt.ylabel("True Positive Rate (TPR)")
plt.title("ROC curve for AVG W2V")
plt.grid()
plt.show()
plt.close()
```



In [0]:

```
#https://stackoverflow.com/questions/35572000/how-can-i-plot-a-confusion-matrix
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
#https://datatofish.com/confusion-matrix-python/

from sklearn.metrics import confusion_matrix as cf_mx

expected_avg_train_w2v = y_train_avg_w2v.values
predicted_avg_train_w2v = final_DT_avg_w2v.predict(x_train_avg_w2v)

expected_avg_test_w2v = y_test_avg_w2v.values
predicted_avg_test_w2v = final_DT_avg_w2v.predict(x_test_avg_w2v)

plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_avg_train_w2v, predicted_avg_train_w2v)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for train data using Avg W2V')

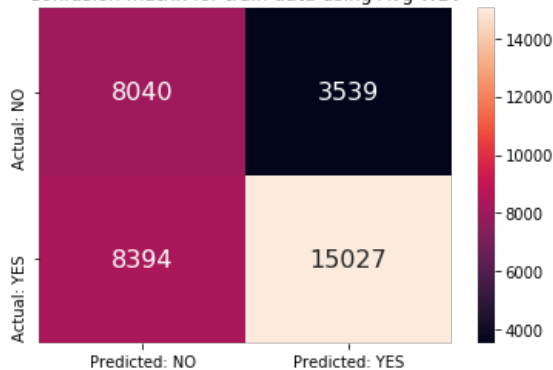
plt.subplot(1,2,2)
cmdf_test=cf_mx(expected_avg_test_w2v, predicted_avg_test_w2v)
df_cm_test = pd.DataFrame(cmdf_test, range(2),range(2))
df_cm_test.columns = ['Predicted: NO','Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for test data using Avg W2V')
```

```

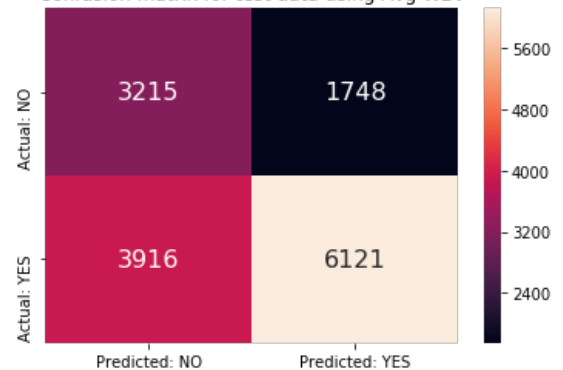
df_cm_test = pd.DataFrame(cm_test, range(2), range(2))
df_cm_test.columns = ['Predicted: NO', 'Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True, annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for test data using Avg W2V')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()

```

Confusion matrix for train data using Avg W2V



Confusion matrix for test data using Avg W2V



## 2.4.4 Applying SGD Classifier brute force on TFIDF W2V, SET 4

### Hyper paramter tuning method: GridSearch

In [0]:

```

#https://www.digitalocean.com/community/tutorials/how-to-plot-data-in-python-3-using-matplotlib
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html
#https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter

from scipy.sparse import hstack
from sklearn.model_selection import GridSearchCV
from sklearn import linear_model
import matplotlib.patches as mpatches
from sklearn.metrics import roc_auc_score

x_train_tfidf_w2v = hstack((categories_one_hot_train, sub_categories_one_hot_train,
grade_cat_one_hot_train,
teacher_prefix_one_hot_train, school_state_one_hot_train,
price_train_standardized,
prev_proj_train_standardized, wc_title_train_standardized,
wc_essay_train_standardized, senti_score_train_standardized,
qty_train_standardized, tfidf_w2v_train_text_vectors,
tfidf_w2v_train_title_vectors))
y_train_tfidf_w2v = df_train['project_is_approved']

x_test_tfidf_w2v = hstack((categories_one_hot_test, sub_categories_one_hot_test,
grade_cat_one_hot_test,
teacher_prefix_one_hot_test, school_state_one_hot_test, price_test_standardized,
prev_proj_test_standardized, wc_title_test_standardized,
wc_essay_test_standardized, senti_score_test_standardized,
qty_test_standardized, tfidf_w2v_test_text_vectors,
tfidf_w2v_test_title_vectors))
y_test_tfidf_w2v = df_test['project_is_approved']

print(x_train_tfidf_w2v.shape, type(x_train_tfidf_w2v), y_train_tfidf_w2v.shape,
type(y_train_tfidf_w2v))
print(x_test_tfidf_w2v.shape, type(x_test_tfidf_w2v), y_test_tfidf_w2v.shape,
type(y_test_tfidf_w2v))

(35000, 206) <class 'scipy.sparse.coo.coo_matrix'> (35000,) <class 'pandas.core.series.Series'>
(15000, 206) <class 'scipy.sparse.coo.coo_matrix'> (15000,) <class 'pandas.core.series.Series'>

```

In [0]:

```

#https://stackabuse.com/cross-validation-and-grid-search-for-model-selection-in-python/
#https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

```

```
#https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
```

```
#Initialising Classifier
```

```
classifier = DecisionTreeClassifier(class_weight='balanced')
```

```
#Brute force approach for finding best K value
```

```
parameters = {'max_depth': [1, 5, 10, 50, 100, 500, 100],  
              'min_samples_split': [5, 10, 100, 500]}
```

```
#Training the model on train data
```

```
DT_tfidf_w2v = GridSearchCV(classifier, parameters, return_train_score=True, cv=3, scoring='roc_auc',  
                             n_jobs=-1)
```

```
DT_tfidf_w2v.fit(x_train_tfidf_w2v, y_train_tfidf_w2v)
```

```
Out[0]:
```

```
GridSearchCV(cv=3, error_score='raise-deprecating',  
             estimator=DecisionTreeClassifier(class_weight='balanced',  
                                              criterion='gini', max_depth=None,  
                                              max_features=None,  
                                              max_leaf_nodes=None,  
                                              min_impurity_decrease=0.0,  
                                              min_impurity_split=None,  
                                              min_samples_leaf=1,  
                                              min_samples_split=2,  
                                              min_weight_fraction_leaf=0.0,  
                                              presort=False, random_state=None,  
                                              splitter='best'),  
             iid='warn', n_jobs=-1,  
             param_grid={'max_depth': [1, 5, 10, 50, 100, 500, 100],  
                         'min_samples_split': [5, 10, 100, 500]},  
             pre_dispatch='2*n_jobs', refit=True, return_train_score=True,  
             scoring='roc_auc', verbose=0)
```

```
In [0]:
```

```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
```

```
#https://stackoverflow.com/questions/20944483/python-3-sort-a-dict-by-its-values/20948781
```

```
print(DT_tfidf_w2v.best_params_) #Gives the best value of parameters from the given range
```

```
print(DT_tfidf_w2v.cv_results_['params'])
```

```
params_train = {}
```

```
params_test = {}
```

```
for i,j in zip(DT_tfidf_w2v.cv_results_['params'],DT_tfidf_w2v.cv_results_['mean_train_score']):  
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))  
    params_train[a]=j
```

```
for i,j in zip(DT_tfidf_w2v.cv_results_['params'],DT_tfidf_w2v.cv_results_['mean_test_score']):  
    a=(str(i['max_depth'])+', '+str(i['min_samples_split']))  
    params_test[a]=j
```

```
params_train = {k: params_train[k] for k in sorted(params_train)}
```

```
params_test = {k: params_test[k] for k in sorted(params_test)}
```

```
print(params_train)
```

```
print(params_test)
```

```
plt.figure(figsize=(22,3))
```

```
plt.plot(params_train.keys(),params_train.values(), label="Train")
```

```
plt.plot(params_test.keys(),params_test.values(), label="Test")
```

```
plt.title('AUC plot for ') 
```

```
plt.xlabel('Parameters')
```

```
plt.ylabel('Area under ROC Curve')
```

```
plt.legend()
```

```
plt.grid()
```

```
plt.show()
```

```
plt.close()
```

```
{'max_depth': 10, 'min_samples_split': 500}
```

```
[{'max_depth': 1, 'min_samples_split': 5}, {'max_depth': 1, 'min_samples_split': 10},
```

```
{ 'max_depth': 1, 'min_samples_split': 100}, {'max_depth': 1, 'min_samples_split': 500},
```

```
{ 'max_depth': 5, 'min_samples_split': 5}, {'max_depth': 5, 'min_samples_split': 10}, {'max_depth':
```

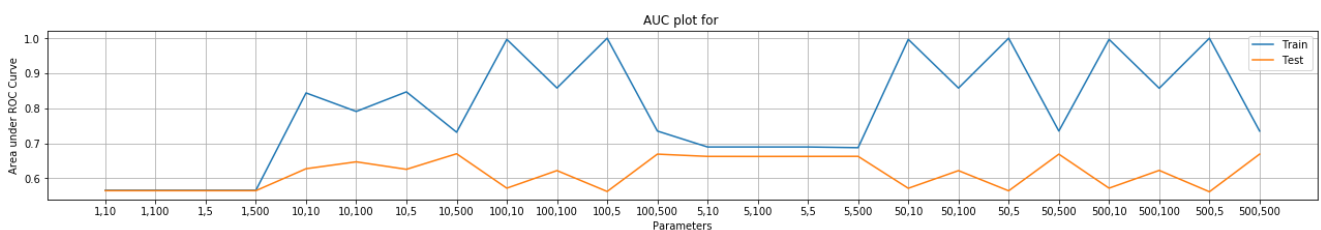
```
5, 'min_samples_split': 100}, {'max_depth': 5, 'min_samples_split': 500}, {'max_depth': 10,
```

```
'min_samples_split': 5}, {'max_depth': 10, 'min_samples_split': 10}, {'max_depth': 10,
```

```

min_samples_split': 5}, {'max_depth': 10, 'min_samples_split': 10}, {'max_depth': 10,
'min_samples_split': 100}, {'max_depth': 10, 'min_samples_split': 500}, {'max_depth': 50,
'min_samples_split': 5}, {'max_depth': 50, 'min_samples_split': 10}, {'max_depth': 50,
'min_samples_split': 100}, {'max_depth': 50, 'min_samples_split': 500}, {'max_depth': 100,
'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100,
'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}, {'max_depth': 500,
'min_samples_split': 5}, {'max_depth': 500, 'min_samples_split': 10}, {'max_depth': 500,
'min_samples_split': 100}, {'max_depth': 500, 'min_samples_split': 500}, {'max_depth': 100,
'min_samples_split': 5}, {'max_depth': 100, 'min_samples_split': 10}, {'max_depth': 100,
'min_samples_split': 100}, {'max_depth': 100, 'min_samples_split': 500}}
{'1,10': 0.565929483669042, '1,100': 0.565929483669042, '1,5': 0.565929483669042, '1,500':
0.565929483669042, '10,10': 0.8437381364028337, '10,100': 0.7908831144553478, '10,5':
0.8467827366618513, '10,500': 0.7315676892651582, '100,10': 0.9965604594257451, '100,100':
0.8578544603782844, '100,5': 0.9998484011274114, '100,500': 0.734987384040488, '5,10':
0.6895634021482019, '5,100': 0.6895634021482019, '5,5': 0.6895634021482019, '5,500':
0.6875124095972517, '50,10': 0.996532630105268, '50,100': 0.8575862076268109, '50,5':
0.9998507462129066, '50,500': 0.7349170967899864, '500,10': 0.9964903715711168, '500,100':
0.8571111891980593, '500,5': 0.9998504529566015, '500,500': 0.7349158937147534}
{'1,10': 0.5648583274471642, '1,100': 0.5648583274471642, '1,5': 0.5648583274471642, '1,500': 0.56
48583274471642, '10,10': 0.6274638181520927, '10,100': 0.6471572174662845, '10,5':
0.6257561364944147, '10,500': 0.6701391882144381, '100,10': 0.5719837137185654, '100,100':
0.6221898736276238, '100,5': 0.5625705509233437, '100,500': 0.6690907913577948, '5,10':
0.6625947756991792, '5,100': 0.662468307925615, '5,5': 0.6625947756991792, '5,500':
0.6627781640188798, '50,10': 0.5716590453179367, '50,100': 0.6220752854380243, '50,5':
0.5644508804752931, '50,500': 0.6690086365292929, '500,10': 0.5720060556812334, '500,100':
0.6227480837507284, '500,5': 0.561753248035648, '500,500': 0.6690830703304224}

```



In [0]:

```

#https://datascience.stackexchange.com/questions/22762/understanding-predict-proba-from-
multioutputclassifier
#https://stackoverflow.com/questions/34894587/should-we-plot-the-roc-curve-for-each-class
#https://stackoverflow.com/questions/55250963/how-to-get-probabilities-for-sgdclassifier-linearsvm

from sklearn.metrics import roc_curve, auc

#training the model on the best K value found in the above result
final_DT_tfidf_w2v = DecisionTreeClassifier(max_depth=10, min_samples_split=500,
class_weight='balanced')
final_DT_tfidf_w2v.fit(x_train_tfidf_w2v, y_train_tfidf_w2v)

x_train_tfidf_w2v_csr=x_train_tfidf_w2v.tocsr()
x_test_tfidf_w2v_csr=x_test_tfidf_w2v.tocsr()

y_train_tfidf_w2v_pred=[]
y_test_tfidf_w2v_pred=[]

#ROC curve function takes the actual values and the predicted probabilities of the positive class
for i in range(0,x_train_tfidf_w2v.shape[0]):
    y_train_tfidf_w2v_pred.extend(final_DT_tfidf_w2v.predict_proba(x_train_tfidf_w2v_csr[i])[:,1])
#[:,1] gives the probability for class 1

for i in range(0,x_test_tfidf_w2v.shape[0]):
    y_test_tfidf_w2v_pred.extend(final_DT_tfidf_w2v.predict_proba(x_test_tfidf_w2v_csr[i])[:,1])

```

In [0]:

```

#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#Calculating FPR and TPR for train and test data
train_tfidf_w2v_fpr, train_tfidf_w2v_tpr, train_tfidf_w2v_thresholds = roc_curve(y_train_tfidf_w2v
, y_train_tfidf_w2v_pred)
test_tfidf_w2v_fpr, test_tfidf_w2v_tpr, test_tfidf_w2v_thresholds = roc_curve(y_test_tfidf_w2v

```

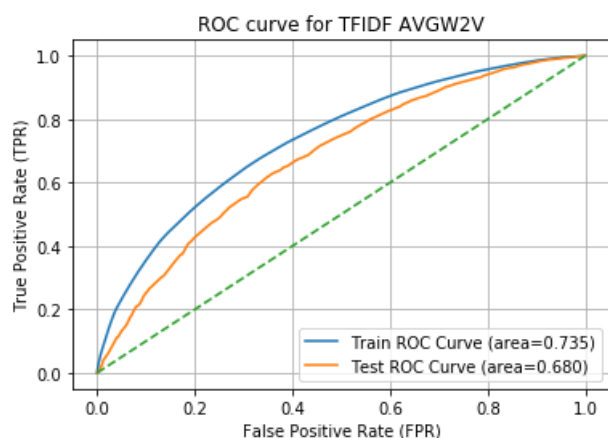
```

test_tfidf_w2v_tpr, test_tfidf_w2v_fpr, test_tfidf_w2v_thresholds = roc_curve(y_test_tfidf_w2v,
y_test_tfidf_w2v_pred)

#Calculating AUC for train and test curves
roc_auc_tfidf_w2v_train=auc(train_tfidf_w2v_fpr,train_tfidf_w2v_tpr)
roc_auc_tfidf_w2v_test=auc(test_tfidf_w2v_fpr,test_tfidf_w2v_tpr)

plt.plot(train_tfidf_w2v_fpr, train_tfidf_w2v_tpr, label="Train ROC Curve (area=%0.3f)" %
roc_auc_tfidf_w2v_train)
plt.plot(test_tfidf_w2v_fpr, test_tfidf_w2v_tpr, label="Test ROC Curve (area=%0.3f)" %
roc_auc_tfidf_w2v_test)
plt.plot([0,1],[0,1],linestyle='--')
plt.legend()
plt.xlabel("False Positive Rate (FPR)")
plt.ylabel("True Positive Rate (TPR)")
plt.title("ROC curve for TFIDF AVGW2V")
plt.grid()
plt.show()
plt.close()

```



In [0]:

```

#https://stackoverflow.com/questions/35572000/how-can-i-plot-a-confusion-matrix
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion\_matrix.html
#https://datatofish.com/confusion-matrix-python/

```

```

from sklearn.metrics import confusion_matrix as cf_mx

expected_tfidf_train_w2v = y_train_tfidf_w2v.values
predicted_tfidf_train_w2v = final_DT_tfidf_w2v.predict(x_train_tfidf_w2v)

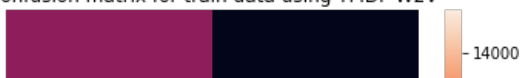
expected_tfidf_test_w2v = y_test_tfidf_w2v.values
predicted_tfidf_test_w2v = final_DT_tfidf_w2v.predict(x_test_tfidf_w2v)

plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_tfidf_train_w2v, predicted_tfidf_train_w2v)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for train data using TFIDF W2V')

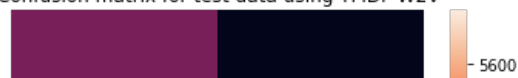
plt.subplot(1,2,2)
cmdf_test=cf_mx(expected_tfidf_test_w2v, predicted_tfidf_test_w2v)
df_cm_test = pd.DataFrame(cmdf_test, range(2),range(2))
df_cm_test.columns = ['Predicted: NO','Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for test data using TFIDF W2V')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()

```

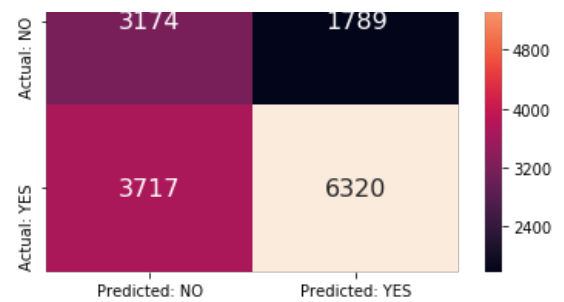
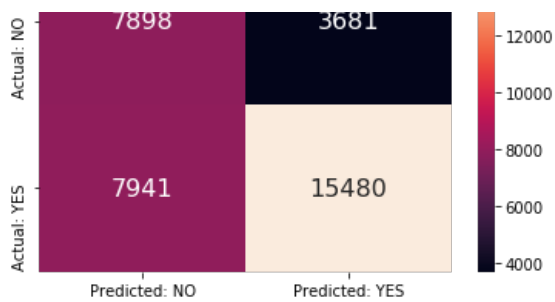
Confusion matrix for train data using TFIDF W2V



Confusion matrix for test data using TFIDF W2V







## 2.5 Select best 5k features from TFIDF , Set 5

### Hyper paramter tuning method: GridSearch

In [0]:

```
#https://www.digitalocean.com/community/tutorials/how-to-plot-data-in-python-3-using-matplotlib
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html
#https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter

from scipy.sparse import hstack
from sklearn.model_selection import GridSearchCV
from sklearn import linear_model
import matplotlib.patches as mpatches
from sklearn.metrics import roc_auc_score

x_train_set5 = hstack((categories_one_hot_train, sub_categories_one_hot_train,
                    grade_cat_one_hot_train,
                    teacher_prefix_one_hot_train, school_state_one_hot_train,
                    price_train_standardized,
                    prev_proj_train_standardized, wc_title_train_standardized,
                    wc_essay_train_standardized, senti_score_train_standardized,
                    qty_train_standardized, text_train_tfidf, title_train_tfidf))
y_train_set5 = df_train['project_is_approved']

x_test_set5 = hstack((categories_one_hot_test, sub_categories_one_hot_test, grade_cat_one_hot_test,
                    teacher_prefix_one_hot_test, school_state_one_hot_test, price_test_standardized,
                    prev_proj_test_standardized, wc_title_test_standardized,
                    wc_essay_test_standardized, senti_score_test_standardized,
                    qty_test_standardized, text_test_tfidf, title_test_tfidf))
y_test_set5 = df_test['project_is_approved']

print(x_train_set5.shape, type(x_train_set5), y_train_set5.shape, type(y_train_set5))
print(x_test_set5.shape, type(x_test_set5), y_test_set5.shape, type(y_test_set5))

(35000, 12176) <class 'scipy.sparse.coo.coo_matrix'> (35000,) <class 'pandas.core.series.Series'>
(15000, 12176) <class 'scipy.sparse.coo.coo_matrix'> (15000,) <class 'pandas.core.series.Series'>
```

In [0]:

```
#Initialising Classifier
classifier = DecisionTreeClassifier(class_weight='balanced')

classifier.fit(x_train_set5, y_train_set5)
```

Out[0]:

```
DecisionTreeClassifier(class_weight='balanced', criterion='gini',
                    max_depth=None, max_features=None, max_leaf_nodes=None,
                    min_impurity_decrease=0.0, min_impurity_split=None,
                    min_samples_leaf=1, min_samples_split=2,
                    min_weight_fraction_leaf=0.0, presort=False,
                    random_state=None, splitter='best')
```

In [0]:

```
#https://datascience.stackexchange.com/questions/31406/tree-decisiontree-feature-importances-numbe
rs-correspond-to-how-features
```

is correspond to how features

```
df_train_set5=pd.DataFrame(x_train_set5.todense())
df_test_set5=pd.DataFrame(x_test_set5.todense())
print(df_train_set5.shape, df_test_set5.shape)

res = dict(zip(df_train_set5.columns, classifier.feature_importances_))

print(res.keys())
print(res.values())

#Deleting all the features with zero importance
delete=[]

for k,v in res.items():
    if v==0:
        delete.append(k)

for i in delete:
    del res[i]

print(len(res))
```

```
(35000, 12176) (15000, 12176)
dict keys([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1715, 1716, 1717, 1718, 1719, 1720, 1721, 1722, 1723, 1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735, 1736, 1737, 1738, 1739, 1740, 1741, 1742, 1743, 1744, 1745, 1746, 1747, 1748, 1749, 1750, 1751, 1752, 1753, 1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768, 1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783, 1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798, 1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843, 1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858, 1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 22
```

1, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1715, 1716, 1717, 1718, 1719, 1720, 1721, 1722, 1723, 1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733,

2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 2680, 2681, 2682, 2683, 2684, 2685, 2686, 2687, 2688, 2689, 2690, 2691, 2692, 2693, 2694, 2695, 2696, 2697, 2698, 2699, 2700, 2701, 2702, 2703, 2704, 2705, 2706, 2707, 2708, 2709, 2710, 2711, 2712, 2713, 2714, 2715, 2716, 2717, 2718, 2719, 2720, 2721, 2722, 2723, 2724, 2725, 2726, 2727, 2728, 2729, 2730, 2731, 2732, 2733, 2734, 2735, 2736, 2737, 2738, 2739, 2740, 2741, 2742, 2743, 2744, 2745, 2746, 2747, 2748, 2749, 2750, 2751, 2752, 2753, 2754, 2755, 2756, 2757, 2758, 2759, 2760, 2761, 2762, 2763, 2764, 2765, 2766, 2767, 2768, 2769, 2770, 2771, 2772, 2773, 2774, 2775, 2776, 2777, 2778, 2779, 2780, 2781, 2782, 2783, 2784, 2785, 2786, 2787, 2788, 2789, 2790, 2791, 2792, 2793, 2794, 2795, 2796, 2797, 2798, 2799, 2800, 2801, 2802, 2803, 2804, 2805, 2806, 2807, 2808, 2809, 2810, 2811, 2812, 2813, 2814, 2815, 2816, 2817, 2818, 2819, 2820, 2821, 2822, 2823, 2824, 2825, 2826, 2827, 2828, 2829, 2830, 2831, 2832, 2833, 2834, 2835, 2836, 2837, 2838, 2839, 2840, 2841, 2842, 2843, 2844, 2845, 2846, 2847, 2848, 2849, 2850, 2851, 2852, 2853, 2854, 2855, 2856, 2857, 2858, 2859, 2860, 2861, 2862, 2863, 2864, 2865, 2866, 2867, 2868, 2869, 2870, 2871, 2872, 2873, 2874, 2875, 2876, 2877, 2878, 2879, 2880, 2881, 2882, 2883, 2884, 2885, 2886, 2887, 2888, 2889, 2890, 2891, 2892, 2893, 2894, 2895, 2896, 2897, 2898, 2899, 2900, 2901, 2902, 2903, 2904, 2905, 2906, 2907, 2908, 2909, 2910, 2911, 2912, 2913, 2914, 2915, 2916, 2917, 2918, 2919, 2920, 2921, 2922, 2923, 2924, 2925, 2926, 2927, 2928, 2929, 2930, 2931, 2932, 2933, 2934, 2935, 2936, 2937, 2938, 2939, 2940, 2941, 2942, 2943, 2944, 2945, 2946, 2947, 2948, 2949, 2950, 2951, 2952, 2953, 2954, 2955, 2956, 2957, 2958, 2959, 2960, 2961, 2962, 2963, 2964, 2965, 2966, 2967, 2968, 2969, 2970, 2971, 2972, 2973, 2974, 2975, 2976, 2977, 2978, 2979, 2980, 2981, 2982, 2983, 2984, 2985, 2986, 2987, 2988, 2989, 2990, 2991, 2992, 2993, 2994, 2995, 2996, 2997, 2998, 2999, 3000, 3001, 3002, 3003, 30

2, 3593, 3594, 3595, 3596, 3597, 3598, 3599, 3600, 3601, 3602, 3603, 3604, 3605, 3606, 3607, 3608, 3609, 3610, 3611, 3612, 3613, 3614, 3615, 3616, 3617, 3618, 3619, 3620, 3621, 3622, 3623, 3624, 3625, 3626, 3627, 3628, 3629, 3630, 3631, 3632, 3633, 3634, 3635, 3636, 3637, 3638, 3639, 3640, 3641, 3642, 3643, 3644, 3645, 3646, 3647, 3648, 3649, 3650, 3651, 3652, 3653, 3654, 3655, 3656, 3657, 3658, 3659, 3660, 3661, 3662, 3663, 3664, 3665, 3666, 3667, 3668, 3669, 3670, 3671, 3672, 3673, 3674, 3675, 3676, 3677, 3678, 3679, 3680, 3681, 3682, 3683, 3684, 3685, 3686, 3687, 3688, 3689, 3690, 3691, 3692, 3693, 3694, 3695, 3696, 3697, 3698, 3699, 3700, 3701, 3702, 3703, 3704, 3705, 3706, 3707, 3708, 3709, 3710, 3711, 3712, 3713, 3714, 3715, 3716, 3717, 3718, 3719, 3720, 3721, 3722, 3723, 3724, 3725, 3726, 3727, 3728, 3729, 3730, 3731, 3732, 3733, 3734, 3735, 3736, 3737, 3738, 3739, 3740, 3741, 3742, 3743, 3744, 3745, 3746, 3747, 3748, 3749, 3750, 3751, 3752, 3753, 3754, 3755, 3756, 3757, 3758, 3759, 3760, 3761, 3762, 3763, 3764, 3765, 3766, 3767, 3768, 3769, 3770, 3771, 3772, 3773, 3774, 3775, 3776, 3777, 3778, 3779, 3780, 3781, 3782, 3783, 3784, 3785, 3786, 3787, 3788, 3789, 3790, 3791, 3792, 3793, 3794, 3795, 3796, 3797, 3798, 3799, 3800, 3801, 3802, 3803, 3804, 3805, 3806, 3807, 3808, 3809, 3810, 3811, 3812, 3813, 3814, 3815, 3816, 3817, 3818, 3819, 3820, 3821, 3822, 3823, 3824, 3825, 3826, 3827, 3828, 3829, 3830, 3831, 3832, 3833, 3834, 3835, 3836, 3837, 3838, 3839, 3840, 3841, 3842, 3843, 3844, 3845, 3846, 3847, 3848, 3849, 3850, 3851, 3852, 3853, 3854, 3855, 3856, 3857, 3858, 3859, 3860, 3861, 3862, 3863, 3864, 3865, 3866, 3867, 3868, 3869, 3870, 3871, 3872, 3873, 3874, 3875, 3876, 3877, 3878, 3879, 3880, 3881, 3882, 3883, 3884, 3885, 3886, 3887, 3888, 3889, 3890, 3891, 3892, 3893, 3894, 3895, 3896, 3897, 3898, 3899, 3900, 3901, 3902, 3903, 3904, 3905, 3906, 3907, 3908, 3909, 3910, 3911, 3912, 3913, 3914, 3915, 3916, 3917, 3918, 3919, 3920, 3921, 3922, 3923, 3924, 3925, 3926, 3927, 3928, 3929, 3930, 3931, 3932, 3933, 3934, 3935, 3936, 3937, 3938, 3939, 3940, 3941, 3942, 3943, 3944, 3945, 3946, 3947, 3948, 3949, 3950, 3951, 3952, 3953, 3954, 3955, 3956, 3957, 3958, 3959, 3960, 3961, 3962, 3963, 3964, 3965, 3966, 3967, 3968, 3969, 3970, 3971, 3972, 3973, 3974, 3975, 3976, 3977, 3978, 3979, 3980, 3981, 3982, 3983, 3984, 3985, 3986, 3987, 3988, 3989, 3990, 3991, 3992, 3993, 3994, 3995, 3996, 3997, 3998, 3999, 4000, 4001, 4002, 4003, 4004, 4005, 4006, 4007, 4008, 4009, 4010, 4011, 4012, 4013, 4014, 4015, 4016, 4017, 4018, 4019, 4020, 4021, 4022, 4023, 4024, 4025, 4026, 4027, 4028, 4029, 4030, 4031, 4032, 4033, 4034, 4035, 4036, 4037, 4038, 4039, 4040, 4041, 4042, 4043, 4044, 4045, 4046, 4047, 4048, 4049, 4050, 4051, 4052, 4053, 4054, 4055, 4056, 4057, 4058, 4059, 4060, 4061, 4062, 4063, 4064, 4065, 4066, 4067, 4068, 4069, 4070, 4071, 4072, 4073, 4074, 4075, 4076, 4077, 4078, 4079, 4080, 4081, 4082, 4083, 4084, 4085, 4086, 4087, 4088, 4089, 4090, 4091, 4092, 4093, 4094, 4095, 4096, 4097, 4098, 4099, 4100, 4101, 4102, 4103, 4104, 4105, 4106, 4107, 4108, 4109, 4110, 4111, 4112, 4113, 4114, 4115, 4116, 4117, 4118, 4119, 4120, 4121, 4122, 4123, 4124, 4125, 4126, 4127, 4128, 4129, 4130, 4131, 4132, 4133, 4134, 4135, 4136, 4137, 4138, 4139, 4140, 4141, 4142, 4143, 4144, 4145, 4146, 4147, 4148, 4149, 4150, 4151, 4152, 4153, 4154, 4155, 4156, 4157, 4158, 4159, 4160, 4161, 4162, 4163, 4164, 4165, 4166, 4167, 4168, 4169, 4170, 4171, 4172, 4173, 4174, 4175, 4176, 4177, 4178, 4179, 4180, 4181, 4182, 4183, 4184, 4185, 4186, 4187, 4188, 4189, 4190, 4191, 4192, 4193, 4194, 4195, 4196, 4197, 4198, 4199, 4200, 4201, 4202, 4203, 4204, 4205, 4206, 4207, 4208, 4209, 4210, 4211, 4212, 4213, 4214, 4215, 4216, 4217, 4218, 4219, 4220, 4221, 4222, 4223, 4224, 4225, 4226, 4227, 4228, 4229, 4230, 4231, 4232, 4233, 4234, 4235, 4236, 4237, 4238, 4239, 4240, 4241, 4242, 4243, 4244, 4245, 4246, 4247, 4248, 4249, 4250, 4251, 4252, 4253, 4254, 4255, 4256, 4257, 4258, 4259, 4260, 4261, 4262, 4263, 4264, 4265, 4266, 4267, 4268, 4269, 4270, 4271, 4272, 4273, 4274,

4863, 4864, 4865, 4866, 4867, 4868, 4869, 4870, 4871, 4872, 4873, 4874, 4875, 4876, 4877, 4878, 4879, 4880, 4881, 4882, 4883, 4884, 4885, 4886, 4887, 4888, 4889, 4890, 4891, 4892, 4893, 4894, 4895, 4896, 4897, 4898, 4899, 4900, 4901, 4902, 4903, 4904, 4905, 4906, 4907, 4908, 4909, 4910, 4911, 4912, 4913, 4914, 4915, 4916, 4917, 4918, 4919, 4920, 4921, 4922, 4923, 4924, 4925, 4926, 4927, 4928, 4929, 4930, 4931, 4932, 4933, 4934, 4935, 4936, 4937, 4938, 4939, 4940, 4941, 4942, 4943, 4944, 4945, 4946, 4947, 4948, 4949, 4950, 4951, 4952, 4953, 4954, 4955, 4956, 4957, 4958, 4959, 4960, 4961, 4962, 4963, 4964, 4965, 4966, 4967, 4968, 4969, 4970, 4971, 4972, 4973, 4974, 4975, 4976, 4977, 4978, 4979, 4980, 4981, 4982, 4983, 4984, 4985, 4986, 4987, 4988, 4989, 4990, 4991, 4992, 4993, 4994, 4995, 4996, 4997, 4998, 4999, 5000, 5001, 5002, 5003, 5004, 5005, 5006, 5007, 5008, 5009, 5010, 5011, 5012, 5013, 5014, 5015, 5016, 5017, 5018, 5019, 5020, 5021, 5022, 5023, 5024, 5025, 5026, 5027, 5028, 5029, 5030, 5031, 5032, 5033, 5034, 5035, 5036, 5037, 5038, 5039, 5040, 5041, 5042, 5043, 5044, 5045, 5046, 5047, 5048, 5049, 5050, 5051, 5052, 5053, 5054, 5055, 5056, 5057, 5058, 5059, 5060, 5061, 5062, 5063, 5064, 5065, 5066, 5067, 5068, 5069, 5070, 5071, 5072, 5073, 5074, 5075, 5076, 5077, 5078, 5079, 5080, 5081, 5082, 5083, 5084, 5085, 5086, 5087, 5088, 5089, 5090, 5091, 5092, 5093, 5094, 5095, 5096, 5097, 5098, 5099, 5100, 5101, 5102, 5103, 5104, 5105, 5106, 5107, 5108, 5109, 5110, 5111, 5112, 5113, 5114, 5115, 5116, 5117, 5118, 5119, 5120, 5121, 5122, 5123, 5124, 5125, 5126, 5127, 5128, 5129, 5130, 5131, 5132, 5133, 5134, 5135, 5136, 5137, 5138, 5139, 5140, 5141, 5142, 5143, 5144, 5145, 5146, 5147, 5148, 5149, 5150, 5151, 5152, 5153, 5154, 5155, 5156, 5157, 5158, 5159, 5160, 5161, 5162, 5163, 5164, 5165, 5166, 5167, 5168, 5169, 5170, 5171, 5172, 5173, 5174, 5175, 5176, 5177, 5178, 5179, 5180, 5181, 5182, 5183, 5184, 5185, 5186, 5187, 5188, 5189, 5190, 5191, 5192, 5193, 5194, 5195, 5196, 5197, 5198, 5199, 5200, 5201, 5202, 5203, 5204, 5205, 5206, 5207, 5208, 5209, 5210, 5211, 5212, 5213, 5214, 5215, 5216, 5217, 5218, 5219, 5220, 5221, 5222, 5223, 5224, 5225, 5226, 5227, 5228, 5229, 5230, 5231, 5232, 5233, 5234, 5235, 5236, 5237, 5238, 5239, 5240, 5241, 5242, 5243, 5244, 5245, 5246, 5247, 5248, 5249, 5250, 5251, 5252, 5253, 5254, 5255, 5256, 5257, 5258, 5259, 5260, 5261, 5262, 5263, 5264, 5265, 5266, 5267, 5268, 5269, 5270, 5271, 5272, 5273, 5274, 5275, 5276, 5277, 5278, 5279, 5280, 5281, 5282, 5283, 5284, 5285, 5286, 5287, 5288, 5289, 5290, 5291, 5292, 5293, 5294, 5295, 5296, 5297, 5298, 5299, 5300, 5301, 5302, 5303, 5304, 5305, 5306, 5307, 5308, 5309, 5310, 5311, 5312, 5313, 5314, 5315, 5316, 5317, 5318, 5319, 5320, 5321, 5322, 5323, 5324, 5325, 5326, 5327, 5328, 5329, 5330, 5331, 5332, 5333, 5334, 5335, 5336, 5337, 5338, 5339, 5340, 5341, 5342, 5343, 5344, 5345, 5346, 5347, 5348, 5349, 5350, 5351, 5352, 5353, 5354, 5355, 5356, 5357, 5358, 5359, 5360, 5361, 5362, 5363, 5364, 5365, 5366, 5367, 5368, 5369, 5370, 5371, 5372, 5373, 5374, 5375, 5376, 5377, 5378, 5379, 5380, 5381, 5382, 5383, 5384, 5385, 5386, 5387, 5388, 5389, 5390, 5391, 5392, 5393, 5394, 5395, 5396, 5397, 5398, 5399, 5400, 5401, 5402, 5403, 5404, 5405, 5406, 5407, 5408, 5409, 5410, 5411, 5412, 5413, 5414, 5415, 5416, 5417, 5418, 5419, 5420, 5421, 5422, 5423, 5424, 5425, 5426, 5427, 5428, 5429, 5430, 5431, 5432, 5433, 5434, 5435, 5436, 5437, 5438, 5439, 5440, 5441, 5442, 5443, 5444, 5445, 5446, 5447, 5448, 5449, 5450, 5451, 5452, 5453, 5454, 5455, 5456, 5457, 5458, 5459, 5460, 5461, 5462, 5463, 5464, 5465, 5466, 5467, 5468, 5469, 5470, 5471, 5472, 5473, 5474, 5475, 5476, 5477, 5478, 5479, 5480, 5481, 5482, 5483, 5484, 5485, 5486, 5487, 5488, 5489, 5490, 5491, 5492, 5493, 5494, 5495, 5496, 5497, 5498, 5499, 5500, 5501, 5502, 5503, 5504, 5505, 5506, 5507, 5508, 5509, 5510, 5511, 5512, 5513, 5514, 5515, 5516, 5517, 5518, 5519, 5520, 5521, 5522, 5523, 5524, 5525, 5526, 5527, 5528, 5529, 5530, 5531, 5532, 5533, 5534, 5535, 5536, 5537, 5538, 5539, 5540, 5541, 5542, 5543, 5544, 55

3, 6134, 6135, 6136, 6137, 6138, 6139, 6140, 6141, 6142, 6143, 6144, 6145, 6146, 6147, 6148, 6149, 6150, 6151, 6152, 6153, 6154, 6155, 6156, 6157, 6158, 6159, 6160, 6161, 6162, 6163, 6164, 6165, 6166, 6167, 6168, 6169, 6170, 6171, 6172, 6173, 6174, 6175, 6176, 6177, 6178, 6179, 6180, 6181, 6182, 6183, 6184, 6185, 6186, 6187, 6188, 6189, 6190, 6191, 6192, 6193, 6194, 6195, 6196, 6197, 6198, 6199, 6200, 6201, 6202, 6203, 6204, 6205, 6206, 6207, 6208, 6209, 6210, 6211, 6212, 6213, 6214, 6215, 6216, 6217, 6218, 6219, 6220, 6221, 6222, 6223, 6224, 6225, 6226, 6227, 6228, 6229, 6230, 6231, 6232, 6233, 6234, 6235, 6236, 6237, 6238, 6239, 6240, 6241, 6242, 6243, 6244, 6245, 6246, 6247, 6248, 6249, 6250, 6251, 6252, 6253, 6254, 6255, 6256, 6257, 6258, 6259, 6260, 6261, 6262, 6263, 6264, 6265, 6266, 6267, 6268, 6269, 6270, 6271, 6272, 6273, 6274, 6275, 6276, 6277, 6278, 6279, 6280, 6281, 6282, 6283, 6284, 6285, 6286, 6287, 6288, 6289, 6290, 6291, 6292, 6293, 6294, 6295, 6296, 6297, 6298, 6299, 6300, 6301, 6302, 6303, 6304, 6305, 6306, 6307, 6308, 6309, 6310, 6311, 6312, 6313, 6314, 6315, 6316, 6317, 6318, 6319, 6320, 6321, 6322, 6323, 6324, 6325, 6326, 6327, 6328, 6329, 6330, 6331, 6332, 6333, 6334, 6335, 6336, 6337, 6338, 6339, 6340, 6341, 6342, 6343, 6344, 6345, 6346, 6347, 6348, 6349, 6350, 6351, 6352, 6353, 6354, 6355, 6356, 6357, 6358, 6359, 6360, 6361, 6362, 6363, 6364, 6365, 6366, 6367, 6368, 6369, 6370, 6371, 6372, 6373, 6374, 6375, 6376, 6377, 6378, 6379, 6380, 6381, 6382, 6383, 6384, 6385, 6386, 6387, 6388, 6389, 6390, 6391, 6392, 6393, 6394, 6395, 6396, 6397, 6398, 6399, 6400, 6401, 6402, 6403, 6404, 6405, 6406, 6407, 6408, 6409, 6410, 6411, 6412, 6413, 6414, 6415, 6416, 6417, 6418, 6419, 6420, 6421, 6422, 6423, 6424, 6425, 6426, 6427, 6428, 6429, 6430, 6431, 6432, 6433, 6434, 6435, 6436, 6437, 6438, 6439, 6440, 6441, 6442, 6443, 6444, 6445, 6446, 6447, 6448, 6449, 6450, 6451, 6452, 6453, 6454, 6455, 6456, 6457, 6458, 6459, 6460, 6461, 6462, 6463, 6464, 6465, 6466, 6467, 6468, 6469, 6470, 6471, 6472, 6473, 6474, 6475, 6476, 6477, 6478, 6479, 6480, 6481, 6482, 6483, 6484, 6485, 6486, 6487, 6488, 6489, 6490, 6491, 6492, 6493, 6494, 6495, 6496, 6497, 6498, 6499, 6500, 6501, 6502, 6503, 6504, 6505, 6506, 6507, 6508, 6509, 6510, 6511, 6512, 6513, 6514, 6515, 6516, 6517, 6518, 6519, 6520, 6521, 6522, 6523, 6524, 6525, 6526, 6527, 6528, 6529, 6530, 6531, 6532, 6533, 6534, 6535, 6536, 6537, 6538, 6539, 6540, 6541, 6542, 6543, 6544, 6545, 6546, 6547, 6548, 6549, 6550, 6551, 6552, 6553, 6554, 6555, 6556, 6557, 6558, 6559, 6560, 6561, 6562, 6563, 6564, 6565, 6566, 6567, 6568, 6569, 6570, 6571, 6572, 6573, 6574, 6575, 6576, 6577, 6578, 6579, 6580, 6581, 6582, 6583, 6584, 6585, 6586, 6587, 6588, 6589, 6590, 6591, 6592, 6593, 6594, 6595, 6596, 6597, 6598, 6599, 6600, 6601, 6602, 6603, 6604, 6605, 6606, 6607, 6608, 6609, 6610, 6611, 6612, 6613, 6614, 6615, 6616, 6617, 6618, 6619, 6620, 6621, 6622, 6623, 6624, 6625, 6626, 6627, 6628, 6629, 6630, 6631, 6632, 6633, 6634, 6635, 6636, 6637, 6638, 6639, 6640, 6641, 6642, 6643, 6644, 6645, 6646, 6647, 6648, 6649, 6650, 6651, 6652, 6653, 6654, 6655, 6656, 6657, 6658, 6659, 6660, 6661, 6662, 6663, 6664, 6665, 6666, 6667, 6668, 6669, 6670, 6671, 6672, 6673, 6674, 6675, 6676, 6677, 6678, 6679, 6680, 6681, 6682, 6683, 6684, 6685, 6686, 6687, 6688, 6689, 6690, 6691, 6692, 6693, 6694, 6695, 6696, 6697, 6698, 6699, 6700, 6701, 6702, 6703, 6704, 6705, 6706, 6707, 6708, 6709, 6710, 6711, 6712, 6713, 6714, 6715, 6716, 6717, 6718, 6719, 6720, 6721, 6722, 6723, 6724, 6725, 6726, 6727, 6728, 6729, 6730, 6731, 6732, 6733, 6734, 6735, 6736, 6737, 6738, 6739, 6740, 6741, 6742, 6743, 6744, 6745, 6746, 6747, 6748, 6749, 6750, 6751, 6752, 6753, 6754, 6755, 6756, 6757, 6758, 6759, 6760, 6761, 6762, 6763, 6764, 6765, 6766, 6767, 6768, 6769, 6770, 6771, 6772, 6773, 6774, 6775, 6776, 6777, 6778, 6779, 6780, 6781, 6782, 6783, 6784, 6785, 6786, 6787, 6788, 6789, 6790, 6791, 6792, 6793, 6794, 6795, 6796, 6797, 6798, 6799, 6800, 6801, 6802, 6803, 6804, 6805, 6806, 6807, 6808, 6809, 6810, 681

7400, 7401, 7402, 7403, 7404, 7405, 7406, 7407, 7408, 7409, 7410, 7411, 7412, 7413, 7414, 7415, 7416, 7417, 7418, 7419, 7420, 7421, 7422, 7423, 7424, 7425, 7426, 7427, 7428, 7429, 7430, 7431, 7432, 7433, 7434, 7435, 7436, 7437, 7438, 7439, 7440, 7441, 7442, 7443, 7444, 7445, 7446, 7447, 7448, 7449, 7450, 7451, 7452, 7453, 7454, 7455, 7456, 7457, 7458, 7459, 7460, 7461, 7462, 7463, 7464, 7465, 7466, 7467, 7468, 7469, 7470, 7471, 7472, 7473, 7474, 7475, 7476, 7477, 7478, 7479, 7480, 7481, 7482, 7483, 7484, 7485, 7486, 7487, 7488, 7489, 7490, 7491, 7492, 7493, 7494, 7495, 7496, 7497, 7498, 7499, 7500, 7501, 7502, 7503, 7504, 7505, 7506, 7507, 7508, 7509, 7510, 7511, 7512, 7513, 7514, 7515, 7516, 7517, 7518, 7519, 7520, 7521, 7522, 7523, 7524, 7525, 7526, 7527, 7528, 7529, 7530, 7531, 7532, 7533, 7534, 7535, 7536, 7537, 7538, 7539, 7540, 7541, 7542, 7543, 7544, 7545, 7546, 7547, 7548, 7549, 7550, 7551, 7552, 7553, 7554, 7555, 7556, 7557, 7558, 7559, 7560, 7561, 7562, 7563, 7564, 7565, 7566, 7567, 7568, 7569, 7570, 7571, 7572, 7573, 7574, 7575, 7576, 7577, 7578, 7579, 7580, 7581, 7582, 7583, 7584, 7585, 7586, 7587, 7588, 7589, 7590, 7591, 7592, 7593, 7594, 7595, 7596, 7597, 7598, 7599, 7600, 7601, 7602, 7603, 7604, 7605, 7606, 7607, 7608, 7609, 7610, 7611, 7612, 7613, 7614, 7615, 7616, 7617, 7618, 7619, 7620, 7621, 7622, 7623, 7624, 7625, 7626, 7627, 7628, 7629, 7630, 7631, 7632, 7633, 7634, 7635, 7636, 7637, 7638, 7639, 7640, 7641, 7642, 7643, 7644, 7645, 7646, 7647, 7648, 7649, 7650, 7651, 7652, 7653, 7654, 7655, 7656, 7657, 7658, 7659, 7660, 7661, 7662, 7663, 7664, 7665, 7666, 7667, 7668, 7669, 7670, 7671, 7672, 7673, 7674, 7675, 7676, 7677, 7678, 7679, 7680, 7681, 7682, 7683, 7684, 7685, 7686, 7687, 7688, 7689, 7690, 7691, 7692, 7693, 7694, 7695, 7696, 7697, 7698, 7699, 7700, 7701, 7702, 7703, 7704, 7705, 7706, 7707, 7708, 7709, 7710, 7711, 7712, 7713, 7714, 7715, 7716, 7717, 7718, 7719, 7720, 7721, 7722, 7723, 7724, 7725, 7726, 7727, 7728, 7729, 7730, 7731, 7732, 7733, 7734, 7735, 7736, 7737, 7738, 7739, 7740, 7741, 7742, 7743, 7744, 7745, 7746, 7747, 7748, 7749, 7750, 7751, 7752, 7753, 7754, 7755, 7756, 7757, 7758, 7759, 7760, 7761, 7762, 7763, 7764, 7765, 7766, 7767, 7768, 7769, 7770, 7771, 7772, 7773, 7774, 7775, 7776, 7777, 7778, 7779, 7780, 7781, 7782, 7783, 7784, 7785, 7786, 7787, 7788, 7789, 7790, 7791, 7792, 7793, 7794, 7795, 7796, 7797, 7798, 7799, 7800, 7801, 7802, 7803, 7804, 7805, 7806, 7807, 7808, 7809, 7810, 7811, 7812, 7813, 7814, 7815, 7816, 7817, 7818, 7819, 7820, 7821, 7822, 7823, 7824, 7825, 7826, 7827, 7828, 7829, 7830, 7831, 7832, 7833, 7834, 7835, 7836, 7837, 7838, 7839, 7840, 7841, 7842, 7843, 7844, 7845, 7846, 7847, 7848, 7849, 7850, 7851, 7852, 7853, 7854, 7855, 7856, 7857, 7858, 7859, 7860, 7861, 7862, 7863, 7864, 7865, 7866, 7867, 7868, 7869, 7870, 7871, 7872, 7873, 7874, 7875, 7876, 7877, 7878, 7879, 7880, 7881, 7882, 7883, 7884, 7885, 7886, 7887, 7888, 7889, 7890, 7891, 7892, 7893, 7894, 7895, 7896, 7897, 7898, 7899, 7900, 7901, 7902, 7903, 7904, 7905, 7906, 7907, 7908, 7909, 7910, 7911, 7912, 7913, 7914, 7915, 7916, 7917, 7918, 7919, 7920, 7921, 7922, 7923, 7924, 7925, 7926, 7927, 7928, 7929, 7930, 7931, 7932, 7933, 7934, 7935, 7936, 7937, 7938, 7939, 7940, 7941, 7942, 7943, 7944, 7945, 7946, 7947, 7948, 7949, 7950, 7951, 7952, 7953, 7954, 7955, 7956, 7957, 7958, 7959, 7960, 7961, 7962, 7963, 7964, 7965, 7966, 7967, 7968, 7969, 7970, 7971, 7972, 7973, 7974, 7975, 7976, 7977, 7978, 7979, 7980, 7981, 7982, 7983, 7984, 7985, 7986, 7987, 7988, 7989, 7990, 7991, 7992, 7993, 7994, 7995, 7996, 7997, 7998, 7999, 8000, 8001, 8002, 8003, 8004, 8005, 8006, 8007, 8008, 8009, 8010, 8011, 8012, 8013, 8014, 8015, 8016, 8017, 8018, 8019, 8020, 8021, 8022, 8023, 8024, 8025, 8026, 8027, 8028, 8029, 8030, 8031, 8032, 8033, 8034, 8035, 8036, 8037, 8038, 8039, 8040, 8041, 8042, 8043, 8044, 8045, 8046, 8047, 8048, 8049, 8050, 8051, 8052, 8053, 8054, 8055, 8056, 8057, 8058, 8059, 8060, 8061, 8062, 8063, 8064, 8065, 8066, 8067, 8068, 8069, 8070, 8071, 8072, 8073, 8074, 8075, 8076, 8077, 8078, 8079, 8080, 8081, 80



4, 8675, 8676, 8677, 8678, 8679, 8680, 8681, 8682, 8683, 8684, 8685, 8686, 8687, 8688, 8689, 8690, 8691, 8692, 8693, 8694, 8695, 8696, 8697, 8698, 8699, 8700, 8701, 8702, 8703, 8704, 8705, 8706, 8707, 8708, 8709, 8710, 8711, 8712, 8713, 8714, 8715, 8716, 8717, 8718, 8719, 8720, 8721, 8722, 8723, 8724, 8725, 8726, 8727, 8728, 8729, 8730, 8731, 8732, 8733, 8734, 8735, 8736, 8737, 8738, 8739, 8740, 8741, 8742, 8743, 8744, 8745, 8746, 8747, 8748, 8749, 8750, 8751, 8752, 8753, 8754, 8755, 8756, 8757, 8758, 8759, 8760, 8761, 8762, 8763, 8764, 8765, 8766, 8767, 8768, 8769, 8770, 8771, 8772, 8773, 8774, 8775, 8776, 8777, 8778, 8779, 8780, 8781, 8782, 8783, 8784, 8785, 8786, 8787, 8788, 8789, 8790, 8791, 8792, 8793, 8794, 8795, 8796, 8797, 8798, 8799, 8800, 8801, 8802, 8803, 8804, 8805, 8806, 8807, 8808, 8809, 8810, 8811, 8812, 8813, 8814, 8815, 8816, 8817, 8818, 8819, 8820, 8821, 8822, 8823, 8824, 8825, 8826, 8827, 8828, 8829, 8830, 8831, 8832, 8833, 8834, 8835, 8836, 8837, 8838, 8839, 8840, 8841, 8842, 8843, 8844, 8845, 8846, 8847, 8848, 8849, 8850, 8851, 8852, 8853, 8854, 8855, 8856, 8857, 8858, 8859, 8860, 8861, 8862, 8863, 8864, 8865, 8866, 8867, 8868, 8869, 8870, 8871, 8872, 8873, 8874, 8875, 8876, 8877, 8878, 8879, 8880, 8881, 8882, 8883, 8884, 8885, 8886, 8887, 8888, 8889, 8890, 8891, 8892, 8893, 8894, 8895, 8896, 8897, 8898, 8899, 8900, 8901, 8902, 8903, 8904, 8905, 8906, 8907, 8908, 8909, 8910, 8911, 8912, 8913, 8914, 8915, 8916, 8917, 8918, 8919, 8920, 8921, 8922, 8923, 8924, 8925, 8926, 8927, 8928, 8929, 8930, 8931, 8932, 8933, 8934, 8935, 8936, 8937, 8938, 8939, 8940, 8941, 8942, 8943, 8944, 8945, 8946, 8947, 8948, 8949, 8950, 8951, 8952, 8953, 8954, 8955, 8956, 8957, 8958, 8959, 8960, 8961, 8962, 8963, 8964, 8965, 8966, 8967, 8968, 8969, 8970, 8971, 8972, 8973, 8974, 8975, 8976, 8977, 8978, 8979, 8980, 8981, 8982, 8983, 8984, 8985, 8986, 8987, 8988, 8989, 8990, 8991, 8992, 8993, 8994, 8995, 8996, 8997, 8998, 8999, 9000, 9001, 9002, 9003, 9004, 9005, 9006, 9007, 9008, 9009, 9010, 9011, 9012, 9013, 9014, 9015, 9016, 9017, 9018, 9019, 9020, 9021, 9022, 9023, 9024, 9025, 9026, 9027, 9028, 9029, 9030, 9031, 9032, 9033, 9034, 9035, 9036, 9037, 9038, 9039, 9040, 9041, 9042, 9043, 9044, 9045, 9046, 9047, 9048, 9049, 9050, 9051, 9052, 9053, 9054, 9055, 9056, 9057, 9058, 9059, 9060, 9061, 9062, 9063, 9064, 9065, 9066, 9067, 9068, 9069, 9070, 9071, 9072, 9073, 9074, 9075, 9076, 9077, 9078, 9079, 9080, 9081, 9082, 9083, 9084, 9085, 9086, 9087, 9088, 9089, 9090, 9091, 9092, 9093, 9094, 9095, 9096, 9097, 9098, 9099, 9100, 9101, 9102, 9103, 9104, 9105, 9106, 9107, 9108, 9109, 9110, 9111, 9112, 9113, 9114, 9115, 9116, 9117, 9118, 9119, 9120, 9121, 9122, 9123, 9124, 9125, 9126, 9127, 9128, 9129, 9130, 9131, 9132, 9133, 9134, 9135, 9136, 9137, 9138, 9139, 9140, 9141, 9142, 9143, 9144, 9145, 9146, 9147, 9148, 9149, 9150, 9151, 9152, 9153, 9154, 9155, 9156, 9157, 9158, 9159, 9160, 9161, 9162, 9163, 9164, 9165, 9166, 9167, 9168, 9169, 9170, 9171, 9172, 9173, 9174, 9175, 9176, 9177, 9178, 9179, 9180, 9181, 9182, 9183, 9184, 9185, 9186, 9187, 9188, 9189, 9190, 9191, 9192, 9193, 9194, 9195, 9196, 9197, 9198, 9199, 9200, 9201, 9202, 9203, 9204, 9205, 9206, 9207, 9208, 9209, 9210, 9211, 9212, 9213, 9214, 9215, 9216, 9217, 9218, 9219, 9220, 9221, 9222, 9223, 9224, 9225, 9226, 9227, 9228, 9229, 9230, 9231, 9232, 9233, 9234, 9235, 9236, 9237, 9238, 9239, 9240, 9241, 9242, 9243, 9244, 9245, 9246, 9247, 9248, 9249, 9250, 9251, 9252, 9253, 9254, 9255, 9256, 9257, 9258, 9259, 9260, 9261, 9262, 9263, 9264, 9265, 9266, 9267, 9268, 9269, 9270, 9271, 9272, 9273, 9274, 9275, 9276, 9277, 9278, 9279, 9280, 9281, 9282, 9283, 9284, 9285, 9286, 9287, 9288, 9289, 9290, 9291, 9292, 9293, 9294, 9295, 9296, 9297, 9298, 9299, 9300, 9301, 9302, 9303, 9304, 9305, 9306, 9307, 9308, 9309, 9310, 9311, 9312, 9313, 9314, 9315, 9316, 9317, 9318, 9319, 9320, 9321, 9322, 9323, 9324, 9325, 9326, 9327, 9328, 9329, 9330, 9331, 9332, 9333, 9334, 9335, 9336, 9337, 9338, 9339, 9340, 9341, 9342, 9343, 9344, 9345, 9346, 9347, 9348, 9349, 9350, 9351, 9352, 9353, 9354, 9355, 9356,



011, 11018, 11019, 11020, 11021, 11022, 11023, 11024, 11025, 11026, 11027, 11028, 11029, 11030, 11  
031, 11032, 11033, 11034, 11035, 11036, 11037, 11038, 11039, 11040, 11041, 11042, 11043, 11044, 11  
045, 11046, 11047, 11048, 11049, 11050, 11051, 11052, 11053, 11054, 11055, 11056, 11057, 11058, 11  
059, 11060, 11061, 11062, 11063, 11064, 11065, 11066, 11067, 11068, 11069, 11070, 11071, 11072, 11  
073, 11074, 11075, 11076, 11077, 11078, 11079, 11080, 11081, 11082, 11083, 11084, 11085, 11086, 11  
087, 11088, 11089, 11090, 11091, 11092, 11093, 11094, 11095, 11096, 11097, 11098, 11099, 11100, 11  
101, 11102, 11103, 11104, 11105, 11106, 11107, 11108, 11109, 11110, 11111, 11112, 11113, 11114, 11  
115, 11116, 11117, 11118, 11119, 11120, 11121, 11122, 11123, 11124, 11125, 11126, 11127, 11128, 11  
129, 11130, 11131, 11132, 11133, 11134, 11135, 11136, 11137, 11138, 11139, 11140, 11141, 11142, 11  
143, 11144, 11145, 11146, 11147, 11148, 11149, 11150, 11151, 11152, 11153, 11154, 11155, 11156, 11  
157, 11158, 11159, 11160, 11161, 11162, 11163, 11164, 11165, 11166, 11167, 11168, 11169, 11170, 11  
171, 11172, 11173, 11174, 11175, 11176, 11177, 11178, 11179, 11180, 11181, 11182, 11183, 11184, 11  
185, 11186, 11187, 11188, 11189, 11190, 11191, 11192, 11193, 11194, 11195, 11196, 11197, 11198, 11  
199, 11200, 11201, 11202, 11203, 11204, 11205, 11206, 11207, 11208, 11209, 11210, 11211, 11212, 11  
213, 11214, 11215, 11216, 11217, 11218, 11219, 11220, 11221, 11222, 11223, 11224, 11225, 11226, 11  
227, 11228, 11229, 11230, 11231, 11232, 11233, 11234, 11235, 11236, 11237, 11238, 11239, 11240, 11  
241, 11242, 11243, 11244, 11245, 11246, 11247, 11248, 11249, 11250, 11251, 11252, 11253, 11254, 11  
255, 11256, 11257, 11258, 11259, 11260, 11261, 11262, 11263, 11264, 11265, 11266, 11267, 11268, 11  
269, 11270, 11271, 11272, 11273, 11274, 11275, 11276, 11277, 11278, 11279, 11280, 11281, 11282, 11  
283, 11284, 11285, 11286, 11287, 11288, 11289, 11290, 11291, 11292, 11293, 11294, 11295, 11296, 11  
297, 11298, 11299, 11300, 11301, 11302, 11303, 11304, 11305, 11306, 11307, 11308, 11309, 11310, 11  
311, 11312, 11313, 11314, 11315, 11316, 11317, 11318, 11319, 11320, 11321, 11322, 11323, 11324, 11  
325, 11326, 11327, 11328, 11329, 11330, 11331, 11332, 11333, 11334, 11335, 11336, 11337, 11338, 11  
339, 11340, 11341, 11342, 11343, 11344, 11345, 11346, 11347, 11348, 11349, 11350, 11351, 11352, 11  
353, 11354, 11355, 11356, 11357, 11358, 11359, 11360, 11361, 11362, 11363, 11364, 11365, 11366, 11  
367, 11368, 11369, 11370, 11371, 11372, 11373, 11374, 11375, 11376, 11377, 11378, 11379, 11380, 11  
381, 11382, 11383, 11384, 11385, 11386, 11387, 11388, 11389, 11390, 11391, 11392, 11393, 11394, 11  
395, 11396, 11397, 11398, 11399, 11400, 11401, 11402, 11403, 11404, 11405, 11406, 11407, 11408, 11  
409, 11410, 11411, 11412, 11413, 11414, 11415, 11416, 11417, 11418, 11419, 11420, 11421, 11422, 11  
423, 11424, 11425, 11426, 11427, 11428, 11429, 11430, 11431, 11432, 11433, 11434, 11435, 11436, 11  
437, 11438, 11439, 11440, 11441, 11442, 11443, 11444, 11445, 11446, 11447, 11448, 11449, 11450, 11  
451, 11452, 11453, 11454, 11455, 11456, 11457, 11458, 11459, 11460, 11461, 11462, 11463, 11464, 11  
465, 11466, 11467, 11468, 11469, 11470, 11471, 11472, 11473, 11474, 11475, 11476, 11477, 11478, 11  
479, 11480, 11481, 11482, 11483, 11484, 11485, 11486, 11487, 11488, 11489, 11490, 11491, 11492, 11  
493, 11494, 11495, 11496, 11497, 11498, 11499, 11500, 11501, 11502, 11503, 11504, 11505, 11506, 11  
507, 11508, 11509, 11510, 11511, 11512, 11513, 11514, 11515, 11516, 11517, 11518, 11519, 11520, 11  
521, 11522, 11523, 11524, 11525, 11526, 11527, 11528, 11529, 11530, 11531, 11532, 11533, 11534, 11  
535, 11536, 11537, 11538, 11539, 11540, 11541, 11542, 11543, 11544, 11545, 11546, 11547, 11548, 11  
549, 11550, 11551, 11552, 11553, 11554, 11555, 11556, 11557, 11558, 11559, 11560, 11561, 11562, 11  
563, 11564, 11565, 11566, 11567, 11568, 11569, 11570, 11571, 11572, 11573, 11574, 11575, 11576, 11  
577, 11578, 11579, 11580, 11581, 11582, 11583, 11584, 11585, 11586, 11587, 11588, 11589, 11590, 11  
591, 11592, 11593, 11594, 11595, 11596, 11597, 11598, 11599, 11600, 11601, 11602, 11603, 11604, 11  
605, 11606, 11607, 11608, 11609, 11610, 11611, 11612, 11613, 11614, 11615, 11616, 11617, 11618, 11  
619, 11620, 11621, 11622, 11623, 11624, 11625, 11626, 11627, 11628, 11629, 11630, 11631, 11632, 11  
633, 11634, 11635, 11636, 11637, 11638, 11639, 11640, 11641, 11642, 11643, 11644, 11645, 11646, 11  
647, 11648, 11649, 11650, 11651, 11652, 11653, 11654, 11655, 11656, 11657, 11658, 11659, 11660, 11  
661, 11662, 11663, 11664, 11665, 11666, 11667, 11668, 11669, 11670, 11671, 11672, 11673, 11674, 11  
675, 11676, 11677, 11678, 11679, 11680, 11681, 11682, 11683, 11684, 11685, 11686, 11687, 11688, 11  
689, 11690, 11691, 11692, 11693, 11694, 11695, 11696, 11697, 11698, 11699, 11700, 11701, 11702, 11  
703, 11704, 11705, 11706, 11707, 11708, 11709, 11710, 11711, 11712, 11713, 11714, 11715, 11716, 11  
717, 11718, 11719, 11720, 11721, 11722, 11723, 11724, 11725, 11726, 11727, 11728, 11729, 11730, 11  
731, 11732, 11733, 11734, 11735, 11736, 11737, 11738, 11739, 11740, 11741, 11742, 11743, 11744, 11  
745, 11746, 11747, 11748, 11749, 11750, 11751, 11752, 11753, 11754, 11755, 11756, 11757, 11758, 11  
759, 11760, 11761, 11762, 11763, 11764, 11765, 11766, 11767, 11768, 11769, 11770, 11771, 11772, 11  
773, 11774, 11775, 11776, 11777, 11778, 11779, 11780, 11781, 11782, 11783, 11784, 11785, 11786, 11  
787, 11788, 11789, 11790, 11791, 11792, 11793, 11794, 11795, 11796, 11797, 11798, 11799, 11800, 11  
801, 11802, 11803, 11804, 11805, 11806, 11807, 11808, 11809, 11810, 11811, 11812, 11813, 11814, 11  
815, 11816, 11817, 11818, 11819, 11820, 11821, 11822, 11823, 11824, 11825, 11826, 11827, 11828, 11  
829, 11830, 11831, 11832, 11833, 11834, 11835, 11836, 11837, 11838, 11839, 11840, 11841, 11842, 11  
843, 11844, 11845, 11846, 11847, 11848, 11849, 11850, 11851, 11852, 11853, 11854, 11855, 11856, 11  
857, 11858, 11859, 11860, 11861, 11862, 11863, 11864, 11865, 11866, 11867, 11868, 11869, 11870, 11  
871, 11872, 11873, 11874, 11875, 11876, 11877, 11878, 11879, 11880, 11881, 11882, 11883, 11884, 11  
885, 11886, 11887, 11888, 11889, 11890, 11891, 11892, 11893, 11894, 11895, 11896, 11897, 11898, 11  
899, 11900, 11901, 11902, 11903, 11904, 11905, 11906, 11907, 11908, 11909, 11910, 11911, 11912, 11  
913, 11914, 11915, 11916, 11917, 11918, 11919, 11920, 11921, 11922, 11923, 11924, 11925, 11926, 11  
927, 11928, 11929, 11930, 11931, 11932, 11933, 11934, 11935, 11936, 11937, 11938, 11939, 11940, 11  
941, 11942, 11943, 11944, 11945, 11946, 11947, 11948, 11949, 11950, 11951, 11952, 11953, 11954, 11  
955, 11956, 11957, 11958, 11959, 11960, 11961, 11962, 11963, 11964, 11965, 11966, 11967, 11968, 11  
969, 11970, 11971, 11972, 11973, 11974, 11975, 11976, 11977, 11978, 11979, 11980, 11981, 11982, 11  
983, 11984, 11985, 11986, 11987, 11988, 11989, 11990, 11991, 11992, 11993, 11994, 11995, 11996, 11  
997, 11998, 11999, 12000, 12001, 12002, 12003, 12004, 12005, 12006, 12007, 12008, 12009, 12010, 12  
011, 12012, 12013, 12014, 12015, 12016, 12017, 12018, 12019, 12020, 12021, 12022, 12023, 12024, 12  
025, 12026, 12027, 12028, 12029, 12030, 12031, 12032, 12033, 12034, 12035, 12036, 12037, 12038, 12  
039, 12040, 12041, 12042, 12043, 12044, 12045, 12046, 12047, 12048, 12049, 12050, 12051, 12052, 12  
053, 12054, 12055, 12056, 12057, 12058, 12059, 12060, 12061, 12062, 12063, 12064, 12065, 12066, 12  
067, 12068, 12069, 12070, 12071, 12072, 12073, 12074, 12075, 12076, 12077, 12078, 12079, 12080, 12  
081, 12082, 12083, 12084, 12085, 12086, 12087, 12088, 12089, 12090, 12091, 12092, 12093, 12094, 12  
097, 12098, 12099, 12100, 12101, 12102, 12103, 12104, 12105, 12106, 12107, 12108, 12109, 12110, 12

```
12096, 12097, 12098, 12099, 12100, 12101, 12102, 12103, 12104, 12105, 12106, 12107, 12108, 12109, 12110, 12111, 12112, 12113, 12114, 12115, 12116, 12117, 12118, 12119, 12120, 12121, 12122, 12123, 12124, 12125, 12126, 12127, 12128, 12129, 12130, 12131, 12132, 12133, 12134, 12135, 12136, 12137, 12138, 12139, 12140, 12141, 12142, 12143, 12144, 12145, 12146, 12147, 12148, 12149, 12150, 12151, 12152, 12153, 12154, 12155, 12156, 12157, 12158, 12159, 12160, 12161, 12162, 12163, 12164, 12165, 12166, 12167, 12168, 12169, 12170, 12171, 12172, 12173, 12174, 12175])
dict_values([0.0002234077252653177, 0.0, 0.00025430345908933593, 0.0006852775403934098, 0.00020176276798995922, 0.00012297702789113333, 0.0, 0.0002833437606488914, 0.0005579231300055108, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015046044746278806, 0.0003590278807648213, 0.0003321260150682853, 8.20145507481764e-05, 0.0, 0.0, 0.0004394888264246298, 6.846853015179927e-05, 7.331199139802604e-05, 0.00015610750055453776, 0.0, 0.0, 7.889287005683438e-05, 0.0004959474395083377, 0.0, 0.0001618937976686592, 0.00013693706030359855, 0.0007075059955101492, 0.0, 0.00014598579125223773, 0.0018837746445602497, 0.0005730955691612572, 0.00020434763438691698, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00042771429381757416, 0.0004175626400239325, 0.0, 0.00014366680793891169, 0.0, 0.0, 0.0002008794706598919, 0.00016705684588138342, 8.429673782712638e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002618725603446462, 0.00015541947716879572, 0.0, 0.0, 0.0, 0.0, 0.00022205056565668174, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00042317933962183763, 0.0, 0.00023832694484708132, 0.0, 0.0, 0.00042171905496248143, 0.0, 0.00023770512521017456, 0.0, 0.00015930067005837426, 0.0, 0.0002496504525470434, 0.0, 0.0, 0.0, 0.0, 7.889287005683432e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.048152624887876624, 0.017202065943698126, 0.0020364126129311364, 0.021203092940509344, 0.0032596799666703485, 0.03302946231257501, 0.0, 0.0, 0.0, 0.00029896662785605974, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016549353681958375, 0.0, 0.0002428583560699986, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00028021710792617747, 0.0, 0.0, 0.0, 0.00016025887972872304, 0.0, 0.0, 0.0, 0.0001707188404616805, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016895252511668428, 0.0, 0.0, 0.0005487443269088512, 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, 0.0, 0.0, 0.0, 0.0, 0.00016164419877495476, 0.0002135925191808987, 0.0, 0.0, 0.0, 0.000361218225703188, 0.0, 0.0, 0.00023481203283554395, 0.0, 0.0, 0.0, 0.0001691580977806235, 0.0, 0.0, 0.0, 0.00015676565836274942, 0.0, 0.0, 0.0, 6.846853015179721e-05, 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, 0.0, 0.0, 0.0, 0.0004288266668105421, 0.0, 0.0, 0.00015176478953942666, 0.0002569095878177277, 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, 0.0, 0.0, 0.0, 0.0006195815567436286, 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, 0.0, 0.0, 0.0, 0.0, 0.00015827807645845722, 0.00012297702789113672, 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, 0.0, 0.0, 0.0, 0.00019771959474231088, 0.0, 0.00030119755971953646, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0004135917631822474, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00013714100794505173, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016578643954550504, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00023117754725731339, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00029378047573267184, 0.0, 0.0, 0.02548790547267409, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000284996512511824, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001668071358336105, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0013631568607869588, 0.0004877165913344975, 0.0002474922308957619, 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, 0.0018092799376241903, 0.0003561594263666829, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.313342881837722e-05, 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, 8.513680160240505e-05, 0.0, 0.0, 0.0, 0.0, 0.00017036619
```

[illegible]

[illegible]

0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0006996770590263516, 0.0, 0.0, 0.0001947560094703549, 0.0, 0.0,  
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0038887874538355107, 0.0, 0.00033526580182002017, 0.0, 0.0, 0.0, 0.0, 8  
.248078639534759e-05, 0.00029018809053020424, 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, 0.0, 0.0, 0.0001530841848339278, 0.00015583371071540536, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,  
, 0.0, 0.0, 0.0001511688425686537, 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, 0.0,  
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.342934329184736e-05, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 7.982216673241913e-05, 0.0  
, 8.322783839778426e-05, 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, 0.0, 0.0, 0.0, 0.0,  
, 0.0,  
7.331199139802604e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 8.375882895674128e-05, 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, 0.0, 0.0, 0.0, 0.0003044474010092792, 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, 0.0, 0.00014801936837289577, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,  
.0, 0.0004902469373265346, 8.007670142730508e-05, 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, 0.0, 0.00016304238772130758, 0.00016625927814690333, 0.0, 0.0004733811465668228,  
0.0001550788027421637, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00020608821909252648, 0.0002515962358948521, 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, 0.0, 0.0, 0.0003077226301710649, 0.0, 0.0, 0.0  
0012297702789112327, 0.0, 0.0, 0.0, 0.0, 0.01439227145632343, 0.0007706890035443536, 0.0, 0.0,  
.000038726155094298255, 0.0, 8.26683462470089e-05, 0.0, 0.0, 0.00014765180286042414,  
0.0005934893382586095, 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, 0.0, 0.0, 0.0, 0.0,  
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.10755247447902e-05, 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, 0.0003422673577965711, 0.0, 0.0019469253774606848, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.3  
61980200682429e-05, 0.000290073215645696, 8.198247569633514e-05, 0.0001229770278911338, 0.0, 0.0,  
0.0, 0.0, 0.0, 0.0, 0.0001440898644131805, 0.00016852939070565554, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.  
.0, 0.00  
038160422054763556, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005252726195705476, 8.503413720872072e-0  
5, 0.0, 0.0001428631726463539, 0.00035069062371492624, 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, 0.0, 0.0, 0.0, 0.00015306533061748226, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,  
00029911712944171505, 0.0, 0.0, 0.0, 0.00022779872554650095, 0.0, 0.0, 0.0, 0.0, 0.0,  
0.0001527313444748711, 0.0008804538165383852, 0.0, 0.0, 0.0, 0.00025669071053778444, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00026226012566702155, 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, 0.0, 0.0, 0.0, 0.0, 0.00014838830812658827, 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, 0.0, 0.0, 0.0, 0.0, 0.0002623113309024963, 0.0, 0  
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00037653012208449966, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,  
0.00016987999514943706, 0.0, 0.0017909927935052883, 0.0, 0.0, 0.0009333406733073599, 0.0,  
0.00020745981926953793, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.259926903110021e-05, 0.0, 0.0, 0.0, 0  
.0, 0.0, 0.0006340065566105375, 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,  
0.0001229770278911391, 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, 0.0, 0.0, 0.0, 0.0,  
0.0, 0.0, 0.0, 0.00023834990936372482, 0.0011981797835243296, 0.0014272447430843526,  
0.0005485660261115257, 0.0, 0.0, 0.0009330649167890092, 0.0, 0.0, 0.0, 0.0, 7.770973858440571e-05,  
0.0, 0.0, 0.0, 0.0, 0.00024407007773813883, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 6.846

```
0.0002179742169238417, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016587902341256745, 0.0,
0.0, 0.0005793902333964865, 0.0, 0.0, 0.0, 6.846853015179721e-05, 0.0, 0.000713912442612681, 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, 0.0, 0.0, 0.00035854692131378303,
0.0, 0.0, 8.215893896844192e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0012548547783582473, 0.0, 0.0, 0.0, 0.00029634657424163174, 8.518800785554428e-05,
0.0009440897279466436, 0.0, 0.0016193289792831753, 0.0, 0.0011302394550375611,
0.00019905181825401056, 0.0005729910981440252, 0.0, 0.00035885179546105665, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0001975960713956115, 0.0, 0.0, 0.0, 0.0, 0.00039611125754160733, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002973978927016906, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
8.24569499271276e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0019916069060288153, 0.0, 0.0004822008283448761, 0.0, 0.0009839022407235212, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0002082615790408135, 0.0, 0.0, 0.0, 0.00020248656012561062, 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, 0.0, 0.0, 0.0, 0.0, 0.00030821801206608193, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00045725854429636486, 0.0, 0.0, 0.0, 0.0, 0.00028456512453227075,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0005781584724791594, 0.00038848889008752824, 0.0013208958205326896, 0.0
005027669108322965, 0.0, 0.0003375638684101895, 0.0, 0.0, 0.0, 0.00014507247402652257,
0.00033057752887417636, 0.00016964935019644204, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 8.09468988343298e-05, 0.0, 0.0, 0.0014042366239470593, 0.0, 0.0, 0.0, 0.0,
0.0009276142362068465, 0.0, 0.0, 0.0, 0.0, 0.00020071005807418242, 0.0,
0.00024421857137914776, 0.0, 0.0004480220766607972, 0.0, 0.0007151767522900529,
7.600011400013806e-05, 0.0, 0.0, 8.505941225489484e-05, 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, 0.0002818174350286698, 0.0006130960214961149, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00038652595182810253, 0.0, 0.0, 0.0, 0.00035135563885169824, 0.0, 0.0002165894120091791, 0.0, 0.0,
0.0, 0.0016614821723890283, 0.0, 0.00029442730206308764, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00031466677319449434, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002364699434451062, 0.0, 0.0, 0.0, 0.0,
0.0, 0.000600755988335496, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002001733334417972,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003514817323922321, 0.0, 0.0, 0.0, 0.0,
0.00015516064100497124, 0.0, 0.0, 0.0008367100545502419, 0.0010566781773155424, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.00026689133914357537, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.009345181028591e-05,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000322572525247951, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0002392695823400591, 0.0, 0.000290543338046886, 0.0, 0.0, 0.0, 0.0, 0.00015353017241147553, 0.0,
0.0, 0.0001645671033609084, 0.0, 0.0, 8.046889901046047e-05, 0.0, 0.0, 0.0, 8.416526647663551e-05,
0.0, 0.0, 0.0, 0.0, 0.0008187124683565177, 0.0, 0.0, 0.0, 0.0, 0.00047327420353269535, 0.0, 0.0, 0.00019625426270395853, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000
3320524186024716, 0.0, 0.0, 0.0, 8.147648711539463e-05, 0.0, 0.00041592786914562105, 0.0, 0.0, 0.0, 0.0,
7.600011400013806e-05, 0.0, 0.0, 7.33119913980256e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
7.600011400013806e-05, 0.0, 0.0, 0.0, 0.00028246558503787137, 0.0005069232943284396, 0.0, 0.0,
0.0, 0.00041984574075651474, 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, 0.0, 0.00
13192739869360573, 0.0004692445551164991, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00022751296967853847, 0.0, 0.00042081819423952084, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0013718749788958022, 0.0, 0.0, 0.0, 0.00029700742985604243, 0.0, 0.0,
0.00043827894167839613, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00047499672868724064, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0001575146275866614, 0.0, 0.0001696137747430149, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0004602587525196957, 0.0, 0.0, 0.0, 0.0, 0.0, 8.413385770832647e-05, 0.0, 0.0,
0.0001543295881415763, 0.0, 0.0, 0.0, 0.00029781781528943066, 0.00023353817487509
```



```
0.0000139489434338784, 0.00016406021237082825, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0004442577189612113, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.001534430469179507, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00023691566114570717, 0.0, 0.0, 0.0, 0.0, 8.36612354475915e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.002850943068156655, 0.0, 0.0, 0.0022036029747052365, 0.0004995896249489157, 0.0, 0.0,
0.0002003606536591852, 0.00022249358426500258, 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, 0.0, 0.0, 0.0, 0.00047991513454877867, 0.0, 0.0, 0.0, 0.0001725910120348179, 0.0, 0.0, 0.0,
0, 0.00020866248495223945, 0.0, 0.0005326827822904123, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0014309408816136145, 7.600011400013806e-05, 0.0, 0.00020674273263494983, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0001346608351058652, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002800238467986596, 0.0, 0.0, 0.0, 0.0,
0.0008770997135230045, 0.0, 0.0, 0.0, 0.0004921563709193739, 0.0, 0.0015378604229951934, 0.0, 0.0,
0.0, 0.0, 0.0008076192027856572, 0.0, 0.0, 0.00015273695200379215, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.137060649571815e-05, 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, 0.0, 0.0, 0.0, 0.0, 7.600011400013934e-05, 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, 0.0, 0.0003211229944620493, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 7.770973858440571e-05, 0.0, 0.0, 0.0, 0.0, 0.0013672662973021, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 0.00015457805578614027, 0.0, 0.0, 0.00016567940039170166,
0.0002060513153912584, 0.0, 0.0, 0.0, 8.182610350500712e-05, 0.0, 0.0014626956683139263, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0007674062447597833, 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,
0.0, 0.0001548555581572765, 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, 0.0, 0.0,
0.0, 0.0, 8.203372813990721e-05, 0.0, 0.0, 0.0, 0.0, 0.00029527820822800927, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.00026018450699573007, 0.0, 0.0, 0.000776081740575751, 0.0008440944594583024, 0.0, 0.0, 0.0,
0.00015578639571949335, 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,
0.00020486620785212977, 0.0, 0.0, 0.0, 0.0, 0.0002907343981044089, 0.00014624025361556535, 0.0, 0.
0, 0.0, 0.0051009881384388245, 8.43211813497654e-05, 0.0, 0.0, 0.0, 0.0012868162530470142,
0.00031197841761852947, 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, 0.0, 0.0, 0.0,
0.0, 0.00212038506041805, 0.0002562764840609695, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00020367826637426558, 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, 0.0,
8.380000805765675e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.510022559708257e-05, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003883852992972221, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0002891159447769801, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000694976899779897, 0.0, 0.0005528142481641466,
0.0002670795460844945, 0.0, 0.0, 0.0, 0.00025039860201934045, 0.0, 0.0, 0.00025008876382854196, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00017086832803955433, 0.0, 0.00021946791790997488, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0008772087613758408, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00029330185286804977, 0.0, 0.
0, 0.0, 0.0, 0.0010143772271599151, 8.490661626351066e-05, 0.0, 0.0, 0.0, 0.0, 0.0,
7.331199139802359e-05, 0.0, 0.0010808073515751469, 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, 0.0, 0.0, 0.00015837757585855872, 0.0, 0.0, 0.00029712959981112853, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.00021365149406662169, 0.0, 0.0, 0.0, 0.0009200153102365917, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00035266110812438416, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.3970053467338e-
05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.140638971939568e-05, 0.0, 0.0, 0.0, 0.0,
0.0003336372448575047, 0.0, 0.0, 0.0002973779735619881, 0.0, 0.0, 0.0, 0.0, 0.0,
```

[illegible]

0.0, 0.0001974325233293974, 0.0, 0.0, 8.118337095799788e-05, 0.0, 0.0, 0.0, 7.770973858440053e-05, 0.0, 0.0, 7.940770604679097e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000155432356824622, 0.0, 0.0, 0.0, 0.0007481901169300848, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001684412616168558, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00020222317787422725, 0.00036977398528385374, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.770973858440566e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002905853205612918, 0.0, 0.0, 0.0, 8.278598670995642e-05, 0.0, 8.125773022259148e-05, 0.0, 0.0, 0.0, 0.0004785974415164254, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00021892959067974975, 0.0, 0.0002529619046366924, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00043341552161286915, 0.0, 0.00030560473592195983, 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, 0.0, 0.0, 0.0, 0.0, 0.00014531690192532107, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 6.846853015179721e-05, 0.0003020891249387957, 0.0, 0.0, 8.432450664754518e-05, 0.0, 8.405298049153776e-05, 0.0, 8.078622006952063e-05, 0.0, 0.0, 0.00012917810431126335, 0.0, 0.000599237773985547, 0.0001618929500338442, 0.0, 0.0, 8.464895522813363e-05, 0.0, 0.00015644356284529802, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.001436400708625175, 0.0, 8.03512064928007e-05, 0.00047746218992909725, 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, 0.0, 0.0, 0.0, 0.0, 0.00024316749798798228, 0.0001627412129915617, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00040004843605470164, 0.0, 0.0, 0.0, 0.00016545125931780098, 0.0, 0.0, 0.00015220236826902743, 0.0, 0.0, 8.168994497405452e-05, 0.0, 0.0, 0.0, 8.227395246168361e-05, 0.0, 0.0, 0.0005665672500505483, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002801651773964964, 0.0, 0.0, 0.0007379020339579158, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001703552450672719, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.01243410915485557, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003625704690939157, 0.0, 0.0006168067696894923, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.405298049153776e-05, 0.0, 0.00030478135753561056, 0.0, 0.0, 0.0, 7.600011400013474e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00023383864468982238, 0.0, 0.0, 0.0, 0.0, 0.0, 0.005145522979907803, 0.0013081993354689483, 0.0, 0.0, 0.0, 0.0, 0.0016406565840711665, 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, 0.0, 0.0, 0.0, 6.846853015179944e-05, 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, 0.0, 0.0, 0.0, 0.0, 0.0002461966080769925, 0.0, 0.0, 0.0, 6.846853015179721e-05, 0.0, 0.0, 0.0, 0.00017750071244240694, 0.0, 0.002375315773425253, 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, 0.0, 0.0, 0.0, 0.00024368496929805085, 0.00020154445199669063, 0.0, 0.0, 0.0002695769200049524, 0.0, 0.0, 0.0, 0.0, 0.0, 8.489380441979616e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000657739084667834, 0.0, 0.00013400463273152607, 0.0, 0.0, 0.00030197588761257546, 0.0001513144126272657, 0.0, 0.0, 0.005386445973921349, 0.0, 0.0, 0.0, 0.0001526861320520883, 0.0, 7.331199139802791e-05, 0.0, 7.911086421686656e-05, 0.0, 0.00028008568780776903, 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, 0.0, 0.0, 0.0, 0.0001972199260935576, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002208949214147394, 0.0, 0.00012297702789114339, 0.0, 0.0, 0.0, 0.0005669641533345916, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.058107104644255e-05, 0.0, 0.0, 0.00024109126768371685, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.462658033252391e-05, 0.0, 0.0, 0.00024516913927074435, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.462658033252391e-05, 0.0, 0.0, 0.0002631404532296897, 0.0, 0.0, 0.0, 0.0, 7.600011400015004e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0012297702789113656, 0.0, 0.0, 0.00015868280848939804, 0.000306985779958085, 0.0003214739179424014, 0.0004038348453459146, 0.0, 0.0, 0.00044141951185934806, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003200763834702105, 0.0, 0.0, 0.0, 0.0, 0.0, 8.231314148787439e-05, 0.0, 0.0, 0.0, 0.0, 0.0003500770090028858, 0.0, 0.0, 0.0, 0.0, 0.0, 8.237875225424642e-05, 0.00173980984151007, 0.00017537410331637168, 0.0, 0.0, 0.00047440421914238376, 0.0003394291800284339, 0.0, 0.00035031471316851595, 0.0, 0.0, 0.0, 0.00020425894048998742, 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, 0.0, 0.0, 0.0, 0.0, 0.0004731023882671973, 0.0010534061697119153, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.11455756923527e-05, 0.0, 0.00040743967034264164, 0.00015334116012155475, 0.0002598673197179284, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.001195195626575077, 0.0047844216441887173, 0.0, 8.103985316192661e-05, 0.00019792991352915367, 0.0, 0.0, 0.0004012797950215517, 0.0004645944913132265, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003550680863383991, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0007649642644587983, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001583295558953932, 0.00014800300076550939, 0.0, 0.0, 0.0001473664162648981, 0.0, 8.24560429288639e-05, 0.0001559472651557385, 0.0, 0.00014933129917942326, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003438315863057534, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00036088475086882747, 0.0, 0.0, 0.0, 0.0, 0.00031516817361662433, 7.404887067303097e-05, 0.00042605369303283725, 0.0, 6.846853015179881e-05, 0.0011076781042664698, 0.0, 0.0, 0.0002772678445354081, 0.0001618937976686592, 0.00023916899926305544, 0.0, 0.0007423082829735995, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005424535495152775, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002634658938332709, 0.0, 0.0, 0.0, 0.0, 0.00019369177299842722, 7.911086421686598e-05, 0.00047941444337112074, 0.0, 0.0, 0.0, 0.0002901710929581887, 0.0, 0.0, 0.0, 0.0, 8.347525626060395e-05, 0.0014594988595279608, 0.0, 0.0, 0.0, 0.00015811469959225603, 0.0, 0.0, 0.0, 8.439711766831446e-05, 0.0009259487696683181, 0.0, 8.292018349429108e-05, 0.0, 0.0, 0.0004831224215443657, 0.0, 0.0, 0.0, 0.0004368054199329806, 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, 0.0, 0.0, 0.0, 0.00020678414416029837, 0.0002889932728163796, 0.0, 0.0, 0.0, 0.0, 0.0002743268092796863, 0.0007931945976108217, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00017106297132261808, 0.0004771244108702473, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00024206999667374563, 0.0, 0.00024683885238024043, 0.0, 0.0, 0.0, 0.0003117150364841291, 0.0, 0.0, 0.000268068271361326, 0

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



◀ ▶

In [0]:

```
print(type(res.keys()), res.keys(), '\n', x_train set5.shape)
```

```
print(loc)
```

```
df_test.set5=df_test.set5.iloc[:, loc]
```

```
<class 'dict_keys'> dict_keys([0, 2, 3, 4, 5, 7, 8, 16, 17, 18, 19, 22, 23, 24, 25, 28, 29, 31, 32, 33, 35, 36, 37, 38, 49, 50, 52, 55, 56, 57, 62, 63, 68, 76, 78, 81, 83, 85, 87, 92, 100, 101, 102, 103, 104, 105, 109, 117, 119, 125, 128, 132, 143, 146, 160, 161, 165, 168, 172, 175, 180, 196, 199, 200, 215, 231, 232, 244, 246, 251, 271, 277, 283, 290, 292, 302, 307, 315, 316, 317, 328, 329, 337, 350, 354, 355, 358, 363, 364, 371, 372, 373, 376, 382, 393, 398, 402, 405, 406, 410, 419, 423, 424, 427, 428, 429, 430, 432, 474, 477, 481, 486, 512, 516, 518, 525, 526, 530, 531, 541, 543, 548, 549, 554, 568, 575, 590, 594, 595, 598, 604, 605, 606, 612, 617, 618, 622, 629, 630, 636, 646, 665, 679, 700, 706, 711, 718, 734, 737, 739, 740, 746, 751, 752, 756, 757, 762, 775, 776, 792, 806, 815, 817, 818, 820, 825, 826, 828, 835, 852, 854, 871, 881, 899, 903, 904, 906, 908, 912, 925, 935, 944, 945, 953, 954, 959, 971, 973, 979, 983, 988, 989, 992, 1005, 1008, 1012, 1016, 1018, 1019, 1028, 1033, 1036, 1038, 1042, 1064, 1067, 1069, 1076, 1079, 1081, 1082, 1084, 1085, 1097, 1100, 1108, 1109, 1116, 1141, 1142, 1145, 1152, 1154, 1157, 1163, 1164, 1168, 1177, 1179, 1194, 1198, 1209, 1215, 1216, 1224, 1225, 1243, 1245, 1250, 1252, 1259, 1266, 1273, 1280, 1291, 1294, 1296, 1297, 1313, 1325, 1334, 1343, 1347, 1351, 1356, 1361, 1365, 1366, 1369, 1376, 1378, 1382, 1389, 1395, 1400, 1402, 1404, 1411, 1421, 1424, 1427, 1442, 1461, 1472, 1475, 1476, 1489, 1498, 1521, 1522, 1529, 1530, 1539, 1550, 1563, 1564, 1566, 1574, 1576, 1577, 1579, 1588, 1645, 1650, 1655, 1656, 1662, 1666, 1667, 1668, 1687, 1690, 1691, 1692, 1693, 1698, 1699, 1700, 1701, 1705, 1711, 1719, 1734, 1735, 1746, 1753, 1759, 1767, 1776, 1778, 1779, 1789, 1790, 1793, 1795, 1806, 1807, 1808, 1820, 1825, 1831, 1832, 1849, 1851, 1854, 1860, 1861, 1862, 1865, 1872, 1894, 1903, 1907, 1910, 1914, 1916, 1928, 1929, 1940, 1943, 1945, 1946, 1953, 1957, 1961, 1965, 1969, 1977, 1984, 1987, 1989, 1993, 1996, 1998, 1999, 2009, 2016, 2019, 2020, 2024, 2025, 2052, 2054, 2057, 2065, 2067, 2069, 2079, 2085, 2086, 2094, 2095, 2097, 2098, 2106, 2108, 2110, 2113, 2115, 2120, 2130, 2136, 2145, 2146, 2152, 2165, 2168, 2198, 2201, 2207, 2210, 2214, 2221, 2226, 2232, 2243, 2257, 2262, 2268, 2273, 2282, 2286, 2291, 2295, 2298, 2304, 2309, 2312, 2339, 2343, 2347, 2349, 2357, 2360, 2362, 2376, 2377, 2412, 2416, 2419, 2420, 2422, 2425, 2444, 2451, 2453, 2461, 2484, 2485, 2486, 2488, 2492, 2498, 2499, 2503, 2504, 2507, 2541, 2545, 2560, 2563, 2571, 2573, 2578, 2579, 2593, 2594, 2605, 2631, 2649, 2651, 2684, 2690, 2706, 2720, 2729, 2730, 2743, 2744, 2746, 2747, 2753, 2754, 2768, 2771, 2775, 2776, 2779, 2781, 2784, 2785, 2808, 2820, 2822, 2829, 2830, 2831, 2832, 2839, 2840, 2866, 2874, 2875, 2877, 2878, 2895, 2903, 2907, 2912, 2913, 2917, 2934, 2950, 2966, 2974, 2982, 2984, 2987, 2989, 2997, 3003, 3013, 3032, 3033, 3034, 3035, 3038, 3043, 3048, 3056, 3058, 3074, 3077, 3079, 3089, 3099, 3101, 3102, 3104, 3106, 3109, 3120, 3125, 3127, 3145, 3152, 3162, 3171, 3176, 3179, 3180, 3194, 3221, 3222, 3231, 3233, 3234, 3241, 3243, 3246, 3249, 3283, 3284, 3289, 3316, 3318, 3319, 3342, 3350, 3354, 3359, 3367, 3375, 3379, 3381, 3382, 3383, 3385, 3388, 3389, 3392, 3395, 3401, 3402, 3403, 3410, 3412, 3413, 3414, 3415, 3416, 3417, 3418, 3419, 3420, 3421, 3422, 3423, 3424, 3425, 3426, 3427, 3428, 3429, 3430, 3431, 3432, 3433, 3434, 3435, 3436, 3437, 3438, 3439, 3440, 3441, 3442, 3443, 3444, 3445, 3446, 3447, 3448, 3449, 3450, 3451, 3452, 3453, 3454, 3455, 3456, 3457, 3458, 3459, 3460, 3461, 3462, 3463, 3464, 3465, 3466, 3467, 3468, 3469, 3470, 3471, 3472, 3473, 3474, 3475, 3476, 3477, 3478, 3479, 3480, 3481, 3482, 3483, 3484, 3485, 3486, 3487, 3488, 3489, 3490, 3491, 3492, 3493, 3494, 3495, 3496, 3497, 3498, 3499, 3500, 3501, 3502, 3503, 3504, 3505, 3506, 3507, 3508, 3509, 3510, 3511, 3512, 3513, 3514, 3515, 3516, 3517, 3518, 3519, 3520, 3521, 3522, 3523, 3524, 3525, 3526, 3527, 3528, 3529, 3530, 3531, 3532, 3533, 3534, 3535, 3536, 3537, 3538, 3539, 3540, 3541, 3542, 3543, 3544, 3545, 3546, 3547, 3548, 3549, 3550, 3551, 3552, 3553, 3554, 3555, 3
```

7, 3425, 3429, 3430, 3432, 3461, 3462, 3464, 3470, 3477, 3479, 3487, 3490, 3500, 3503, 3506, 3508, 3526, 3529, 3539, 3543, 3544, 3545, 3547, 3549, 3550, 3551, 3553, 3563, 3568, 3576, 3584, 3597, 3599, 3601, 3607, 3611, 3631, 3640, 3645, 3651, 3652, 3653, 3654, 3656, 3660, 3661, 3662, 3674, 3677, 3682, 3688, 3690, 3692, 3694, 3695, 3698, 3711, 3712, 3717, 3721, 3723, 3726, 3728, 3734, 3742, 3747, 3758, 3766, 3771, 3774, 3775, 3782, 3791, 3799, 3805, 3807, 3812, 3815, 3818, 3822, 3827, 3832, 3835, 3850, 3854, 3856, 3860, 3863, 3873, 3877, 3878, 3882, 3896, 3897, 3905, 3907, 3919, 3923, 3926, 3932, 3938, 3940, 3949, 3955, 3958, 3962, 3963, 3966, 3973, 3979, 3980, 3981, 3989, 3991, 3992, 3994, 3999, 4000, 4004, 4006, 4011, 4014, 4016, 4017, 4022, 4023, 4026, 4027, 4034, 4060, 4066, 4069, 4072, 4074, 4075, 4076, 4083, 4085, 4087, 4095, 4101, 4109, 4111, 4114, 4116, 4121, 4129, 4133, 4137, 4150, 4152, 4153, 4158, 4168, 4177, 4178, 4183, 4193, 4194, 4202, 4217, 4233, 4236, 4244, 4257, 4265, 4266, 4268, 4271, 4272, 4279, 4280, 4281, 4284, 4296, 4306, 4307, 4308, 4314, 4322, 4328, 4332, 4341, 4344, 4345, 4348, 4349, 4365, 4369, 4373, 4375, 4387, 4388, 4390, 4398, 4417, 4422, 4426, 4428, 4433, 4436, 4456, 4474, 4485, 4495, 4500, 4508, 4511, 4512, 4516, 4518, 4524, 4538, 4555, 4560, 4566, 4569, 4570, 4574, 4587, 4592, 4593, 4597, 4598, 4602, 4603, 4620, 4621, 4629, 4642, 4653, 4662, 4668, 4674, 4676, 4677, 4681, 4684, 4693, 4695, 4706, 4727, 4732, 4733, 4739, 4741, 4756, 4759, 4766, 4770, 4776, 4788, 4798, 4803, 4806, 4839, 4844, 4850, 4855, 4863, 4871, 4878, 4882, 4886, 4890, 4892, 4901, 4931, 4932, 4934, 4935, 4937, 4943, 4950, 4951, 4953, 4956, 4959, 4960, 4961, 4976, 4978, 4981, 4982, 5015, 5021, 5028, 5030, 5041, 5044, 5046, 5050, 5055, 5058, 5065, 5070, 5082, 5084, 5085, 5095, 5106, 5107, 5114, 5116, 5121, 5122, 5123, 5124, 5128, 5129, 5157, 5158, 5159, 5160, 5169, 5194, 5214, 5222, 5241, 5242, 5248, 5249, 5270, 5275, 5306, 5307, 5309, 5327, 5328, 5347, 5371, 5384, 5386, 5395, 5397, 5399, 5405, 5407, 5411, 5414, 5417, 5424, 5425, 5427, 5439, 5440, 5442, 5452, 5460, 5462, 5477, 5479, 5481, 5490, 5491, 5495, 5496, 5497, 5499, 5503, 5509, 5511, 5529, 5539, 5552, 5553, 5555, 5556, 5559, 5560, 5562, 5566, 5568, 5579, 5593, 5597, 5598, 5599, 5600, 5601, 5604, 5605, 5607, 5608, 5619, 5625, 5627, 5628, 5629, 5630, 5633, 5641, 5644, 5651, 5674, 5678, 5682, 5687, 5688, 5692, 5696, 5698, 5703, 5704, 5706, 5708, 5717, 5730, 5734, 5736, 5737, 5748, 5749, 5751, 5757, 5761, 5767, 5769, 5770, 5771, 5781, 5782, 5787, 5792, 5795, 5798, 5801, 5808, 5810, 5811, 5816, 5821, 5827, 5829, 5835, 5839, 5841, 5852, 5862, 5868, 5876, 5879, 5892, 5922, 5926, 5927, 5929, 5936, 5937, 5940, 5944, 5949, 5953, 5959, 5960, 5962, 5965, 5972, 5981, 5991, 5992, 5993, 5994, 5996, 5997, 6000, 6001, 6002, 6006, 6016, 6020, 6021, 6026, 6027, 6037, 6043, 6044, 6052, 6063, 6064, 6067, 6070, 6074, 6077, 6081, 6084, 6095, 6099, 6111, 6122, 6123, 6132, 6140, 6143, 6173, 6177, 6184, 6186, 6192, 6194, 6208, 6217, 6218, 6221, 6223, 6225, 6228, 6230, 6231, 6234, 6236, 6248, 6250, 6251, 6263, 6264, 6269, 6273, 6276, 6279, 6283, 6286, 6291, 6294, 6302, 6309, 6325, 6327, 6335, 6337, 6341, 6347, 6353, 6354, 6359, 6370, 6383, 6387, 6391, 6393, 6408, 6409, 6412, 6418, 6429, 6431, 6434, 6454, 6457, 6461, 6463, 6465, 6467, 6480, 6488, 6490, 6494, 6502, 6505, 6513, 6522, 6525, 6530, 6537, 6540, 6563, 6564, 6565, 6568, 6576, 6582, 6587, 6593, 6594, 6595, 6598, 6599, 6601, 6605, 6619, 6620, 6628, 6630, 6631, 6632, 6638, 6656, 6658, 6674, 6677, 6678, 6688, 6698, 6704, 6726, 6729, 6740, 6755, 6757, 6760, 6762, 6768, 6773, 6774, 6775, 6777, 6791, 6794, 6811, 6812, 6814, 6819, 6824, 6829, 6830, 6831, 6835, 6840, 6841, 6845, 6849, 6850, 6852, 6855, 6859, 6870, 6881, 6885, 6901, 6907, 6908, 6914, 6916, 6920, 6923, 6936, 6937, 6938, 6949, 6964, 6995, 6999, 7005, 7014, 7020, 7021, 7027, 7035, 7036, 7038, 7073, 7076, 7081, 7092, 7098, 7103, 7108, 7113, 7114, 7117, 7146, 7147, 7154, 7162, 7184, 7185, 7201, 7209, 7210, 7216, 7218, 7220, 7228, 7230, 7232, 7241, 7244, 7249, 7251, 7253, 7287, 7290, 7291, 7293, 7294, 7302, 7303, 7308, 7310, 7320, 7329, 7337, 7338, 7344, 7345, 7349, 7356, 7357, 7369, 7371, 7379, 7381, 7383, 7389, 7396, 7402, 7407, 7425, 7430, 7431, 7432, 7440, 7446, 7454, 7457, 7479, 7487, 7490, 7495, 7500, 7503, 7504, 7507, 7508, 7520, 7530, 7535, 7548, 7562, 7567, 7573, 7575, 7583, 7586, 7587, 7590, 7599, 7604, 7610, 7619, 7631, 7656, 7687, 7694, 7698, 7701, 7702, 7704, 7705, 7709, 7711, 7715, 7720, 7723, 7729, 7731, 7735, 7736, 7760, 7770, 7784, 7791, 7801, 7804, 7824, 7831, 7834, 7839, 7850, 7856, 7880, 7900, 7907, 7911, 7922, 7931, 7944, 7948, 7961, 7962, 7972, 7973, 7975, 7976, 7985, 8000, 8015, 8024, 8028, 8029, 8031, 8032, 8050, 8053, 8083, 8098, 8101, 8106, 8131, 8135, 8161, 8162, 8176, 8178, 8186, 8187, 8192, 8198, 8224, 8234, 8249, 8254, 8256, 8258, 8259, 8277, 8293, 8299, 8303, 8304, 8334, 8336, 8340, 8343, 8344, 8347, 8348, 8349, 8352, 8361, 8374, 8378, 8400, 8402, 8405, 8406, 8420, 8425, 8427, 8429, 8433, 8438, 8443, 8448, 8459, 8464, 8470, 8474, 8482, 8486, 8488, 8489, 8495, 8496, 8497, 8499, 8508, 8512, 8530, 8532, 8534, 8536, 8537, 8542, 8544, 8548, 8560, 8569, 8574, 8582, 8583, 8593, 8594, 8601, 8608, 8610, 8622, 8628, 8651, 8662, 8664, 8667, 8670, 8671, 8672, 8677, 8685, 8687, 8688, 8689, 8725, 8749, 8759, 8766, 8770, 8773, 8775, 8782, 8792, 8794, 8796, 8798, 8803, 8807, 8811, 8813, 8820, 8842, 8843, 8844, 8849, 8860, 8867, 8872, 8882, 8883, 8889, 8891, 8896, 8898, 8900, 8934, 8936, 8938, 8957, 8963, 8971, 8974, 8979, 8991, 8993, 8996, 8997, 9008, 9009, 9016, 9019, 9021, 9025, 9028, 9035, 9037, 9042, 9046, 9050, 9066, 9068, 9072, 9073, 9077, 9093, 9094, 9098, 9099, 9101, 9102, 9105, 9113, 9133, 9134, 9138, 9139, 9144, 9163, 9167, 9168, 9175, 9176, 9177, 9179, 9180, 9182, 9183, 9189, 9197, 9198, 9203, 9207, 9229, 9231, 9242, 9248, 9261, 9266, 9273, 9284, 9286, 9287, 9288, 9291, 9292, 9293, 9296, 9310, 9318, 9327, 9353, 9356, 9364, 9370, 9371, 9374, 9376, 9378, 9384, 9387, 9388, 9394, 9397, 9398, 9399, 9404, 9421, 9423, 9434, 9436, 9437, 9438, 9439, 9441, 9444, 9448, 9456, 9458, 9465, 9472, 9474, 9475, 9487, 9508, 9511, 9513, 9514, 9519, 9526, 9527, 9530, 9533, 9535, 9538, 9539, 9540, 9552, 9563, 9564, 9565, 9567, 9568, 9570, 9580, 9581, 9586, 9591, 9600, 9624, 9628, 9634, 9636, 9645, 9646, 9653, 9660, 9665, 9673, 9676, 9687, 9695, 9698, 9705, 9707, 9722, 9727, 9734, 9739, 9743, 9754, 9762, 9765, 9780, 9784, 9803, 9819, 9821, 9851, 9860, 9871, 9872, 9899, 9910, 9922, 9925, 9927, 9930, 9938, 9968, 9970, 9992, 9998, 10005, 10006, 10008, 10010, 10017, 10056, 10061, 10074, 10083, 10088, 10089, 10093, 10094, 10099, 10107, 10110, 10120, 10123, 10125, 10135, 10136, 10139, 10180, 10196, 10197, 10215, 10221, 10235, 10239, 10242, 10273, 10276, 10289, 10290, 10291, 10293, 10297, 10308, 10313, 10317, 10325, 10329, 10330, 10341, 10350, 10357, 10359, 10360, 10361, 10368, 10371, 10373, 10380, 10398, 10400, 10422, 10424, 10439, 10448, 10452, 10457, 10458, 10460, 10465, 10472, 10474, 10482, 10486, 10487, 10488, 10494, 10495, 10500, 10508, 10514, 10515, 10517, 10525, 10536, 10540, 10557, 10563, 10573, 10576, 10579, 10580, 10591, 10621, 10652, 10654, 10655, 10682, 10687, 10689, 10708, 10721, 10750, 10800, 10815, 10827, 10836, 10838, 10840, 10853, 10863, 10996, 10997, 11035, 11050, 11075, 11080, 11088, 11097, 11106, 11132, 11137, 11141, 11154, 11157, 11159, 11161, 11171, 11177, 11187, 11194, 11205, 11212, 11216, 11223, 11244, 11257, 11271, 11302, 11309, 11350, 11359, 11361, 11362, 11376, 11381, 11394, 11395, 11399, 11405, 11409, 11412,

11413, 11422, 11424, 11454, 11459, 11472, 11478, 11483, 11484, 11486, 11557, 11559, 11561, 11569, 11593, 11599, 11605, 11606, 11620, 11625, 11643, 11660, 11677, 11686, 11708, 11741, 11743, 11745, 11748, 11770, 11790, 11801, 11810, 11813, 11816, 11822, 11830, 11835, 11852, 11862, 11905, 11906, 11910, 11916, 11923, 11925, 11926, 11928, 11958, 11959, 11968, 12011, 12035, 12048, 12050, 12054, 12087, 12088, 12122, 12152, 12158, 12161, 12164, 12170])  
(35000, 12176)  
[0, 2, 3, 4, 5, 7, 8, 16, 17, 18, 19, 22, 23, 24, 25, 28, 29, 31, 32, 33, 35, 36, 37, 38, 49, 50, 52, 55, 56, 57, 62, 63, 68, 76, 78, 81, 83, 85, 87, 92, 100, 101, 102, 103, 104, 105, 109, 117, 119, 125, 128, 132, 143, 146, 160, 161, 165, 168, 172, 175, 180, 196, 199, 200, 215, 231, 232, 244, 246, 251, 271, 277, 283, 290, 292, 302, 307, 315, 316, 317, 328, 329, 337, 350, 354, 355, 358, 363, 364, 371, 372, 373, 376, 382, 393, 398, 402, 405, 406, 410, 419, 423, 424, 427, 428, 429, 430, 432, 474, 477, 481, 486, 512, 516, 518, 525, 526, 530, 531, 541, 543, 548, 549, 554, 568, 575, 590, 594, 595, 598, 604, 605, 606, 612, 617, 618, 622, 629, 630, 636, 646, 665, 679, 700, 706, 711, 718, 734, 737, 739, 740, 746, 751, 752, 756, 757, 762, 775, 776, 792, 806, 815, 817, 818, 820, 825, 826, 828, 835, 852, 854, 871, 881, 899, 903, 904, 906, 908, 912, 925, 935, 944, 945, 953, 954, 959, 971, 973, 979, 983, 988, 989, 992, 1005, 1008, 1012, 1016, 1018, 1019, 1028, 1033, 1036, 1038, 1042, 1064, 1067, 1069, 1076, 1079, 1081, 1082, 1084, 1085, 1097, 1100, 1108, 1109, 1116, 1141, 1142, 1145, 1152, 1154, 1157, 1163, 1164, 1168, 1177, 1179, 1194, 1198, 1209, 1215, 1216, 1224, 1225, 1243, 1245, 1250, 1252, 1259, 1266, 1273, 1280, 1291, 1294, 1296, 1297, 1313, 1325, 1334, 1343, 1347, 1351, 1356, 1361, 1365, 1366, 1369, 1376, 1378, 1382, 1389, 1395, 1400, 1402, 1404, 1411, 1421, 1424, 1427, 1442, 1461, 1472, 1475, 1476, 1489, 1498, 1521, 1522, 1529, 1530, 1539, 1550, 1563, 1564, 1566, 1574, 1576, 1577, 1579, 1588, 1645, 1650, 1655, 1656, 1662, 1666, 1667, 1668, 1687, 1690, 1691, 1692, 1693, 1698, 1699, 1700, 1701, 1705, 1711, 1719, 1734, 1735, 1746, 1753, 1759, 1767, 1776, 1778, 1779, 1789, 1790, 1793, 1795, 1806, 1807, 1808, 1820, 1825, 1831, 1832, 1849, 1851, 1854, 1860, 1861, 1862, 1865, 1872, 1894, 1903, 1907, 1910, 1914, 1916, 1928, 1929, 1940, 1943, 1945, 1946, 1953, 1957, 1961, 1965, 1969, 1977, 1984, 1987, 1989, 1993, 1996, 1998, 1999, 2009, 2016, 2019, 2020, 2024, 2025, 2052, 2054, 2057, 2065, 2067, 2069, 2079, 2085, 2086, 2094, 2095, 2097, 2098, 2106, 2108, 2110, 2113, 2115, 2120, 2130, 2136, 2145, 2146, 2152, 2165, 2168, 2198, 2201, 2207, 2210, 2214, 2221, 2226, 2232, 2243, 2257, 2262, 2268, 2273, 2282, 2286, 2291, 2295, 2298, 2304, 2309, 2312, 2339, 2343, 2347, 2349, 2357, 2360, 2362, 2376, 2377, 2412, 2416, 2419, 2420, 2422, 2425, 2444, 2451, 2453, 2461, 2484, 2485, 2486, 2488, 2492, 2498, 2499, 2503, 2504, 2507, 2541, 2545, 2560, 2563, 2571, 2573, 2578, 2579, 2593, 2594, 2605, 2631, 2649, 2651, 2684, 2690, 2706, 2720, 2729, 2730, 2743, 2744, 2746, 2747, 2753, 2754, 2768, 2771, 2775, 2776, 2779, 2781, 2784, 2785, 2808, 2820, 2822, 2829, 2830, 2831, 2832, 2839, 2840, 2866, 2874, 2875, 2877, 2878, 2895, 2903, 2907, 2912, 2913, 2917, 2934, 2950, 2966, 2974, 2982, 2984, 2987, 2989, 2997, 3003, 3013, 3032, 3033, 3034, 3035, 3038, 3043, 3048, 3056, 3058, 3074, 3077, 3079, 3089, 3099, 3101, 3102, 3104, 3106, 3109, 3120, 3125, 3127, 3145, 3152, 3162, 3171, 3176, 3179, 3180, 3194, 3221, 3222, 3231, 3233, 3234, 3241, 3243, 3246, 3249, 3283, 3284, 3289, 3316, 3318, 3319, 3342, 3350, 3354, 3359, 3367, 3375, 3379, 3381, 3382, 3383, 3385, 3388, 3389, 3392, 3395, 3401, 3402, 3403, 3410, 3412, 3417, 3425, 3429, 3430, 3432, 3461, 3462, 3464, 3470, 3477, 3479, 3487, 3490, 3500, 3503, 3506, 3508, 3526, 3529, 3539, 3543, 3544, 3545, 3547, 3549, 3550, 3551, 3553, 3563, 3568, 3576, 3584, 3597, 3599, 3601, 3607, 3611, 3631, 3640, 3645, 3651, 3652, 3653, 3654, 3656, 3660, 3661, 3662, 3674, 3677, 3682, 3688, 3690, 3692, 3694, 3695, 3698, 3711, 3712, 3717, 3721, 3723, 3726, 3728, 3734, 3742, 3747, 3758, 3766, 3771, 3774, 3775, 3782, 3791, 3799, 3805, 3807, 3812, 3815, 3818, 3822, 3827, 3832, 3835, 3850, 3854, 3856, 3860, 3863, 3873, 3877, 3878, 3882, 3896, 3897, 3905, 3907, 3919, 3923, 3926, 3932, 3938, 3940, 3949, 3955, 3958, 3962, 3963, 3966, 3973, 3979, 3980, 3981, 3989, 3991, 3992, 3994, 3999, 4000, 4004, 4006, 4011, 4014, 4016, 4017, 4022, 4023, 4026, 4027, 4034, 4060, 4066, 4069, 4072, 4074, 4075, 4076, 4083, 4085, 4087, 4095, 4101, 4109, 4111, 4114, 4116, 4121, 4129, 4133, 4137, 4150, 4152, 4153, 4158, 4168, 4177, 4178, 4183, 4193, 4194, 4202, 4217, 4233, 4236, 4244, 4257, 4265, 4266, 4268, 4271, 4272, 4279, 4280, 4281, 4284, 4296, 4306, 4307, 4308, 4314, 4322, 4328, 4332, 4341, 4344, 4345, 4348, 4349, 4365, 4369, 4373, 4375, 4387, 4388, 4390, 4398, 4417, 4422, 4426, 4428, 4433, 4436, 4456, 4474, 4485, 4495, 4500, 4508, 4511, 4512, 4516, 4518, 4524, 4538, 4555, 4560, 4566, 4569, 4570, 4574, 4587, 4592, 4593, 4597, 4598, 4602, 4603, 4620, 4621, 4629, 4642, 4653, 4662, 4668, 4674, 4676, 4677, 4681, 4684, 4693, 4695, 4706, 4727, 4732, 4733, 4739, 4741, 4756, 4759, 4766, 4770, 4776, 4788, 4798, 4803, 4806, 4839, 4844, 4850, 4855, 4863, 4871, 4878, 4882, 4886, 4890, 4892, 4901, 4931, 4932, 4934, 4935, 4937, 4943, 4950, 4951, 4953, 4956, 4959, 4960, 4961, 4976, 4978, 4981, 4982, 5015, 5021, 5028, 5030, 5041, 5044, 5046, 5050, 5055, 5058, 5065, 5070, 5082, 5084, 5085, 5095, 5106, 5107, 5114, 5116, 5121, 5122, 5123, 5124, 5128, 5129, 5157, 5158, 5159, 5160, 5169, 5194, 5214, 5222, 5241, 5242, 5248, 5249, 5270, 5275, 5306, 5307, 5309, 5327, 5328, 5347, 5371, 5384, 5386, 5395, 5397, 5399, 5405, 5407, 5411, 5414, 5417, 5424, 5425, 5427, 5439, 5440, 5442, 5452, 5460, 5462, 5477, 5479, 5481, 5490, 5491, 5495, 5495, 5496, 5497, 5499, 5503, 5509, 5511, 5529, 5539, 5552, 5553, 5555, 5556, 5559, 5560, 5562, 5566, 5568, 5579, 5593, 5597, 5598, 5599, 5600, 5601, 5604, 5605, 5607, 5608, 5619, 5625, 5627, 5628, 5629, 5630, 5633, 5641, 5644, 5651, 5674, 5678, 5682, 5687, 5688, 5692, 5696, 5698, 5703, 5704, 5706, 5708, 5717, 5730, 5734, 5736, 5737, 5748, 5749, 5751, 5757, 5761, 5767, 5769, 5770, 5771, 5781, 5782, 5787, 5792, 5795, 5798, 5801, 5808, 5810, 5811, 5816, 5821, 5827, 5829, 5835, 5839, 5841, 5852, 5862, 5868, 5876, 5879, 5892, 5922, 5926, 5927, 5929, 5936, 5937, 5940, 5944, 5949, 5953, 5959, 5960, 5962, 5965, 5972, 5981, 5991, 5992, 5993, 5994, 5996, 5997, 6000, 6001, 6002, 6006, 6016, 6020, 6021, 6026, 6027, 6037, 6043, 6044, 6052, 6063, 6064, 6067, 6070, 6074, 6077, 6081, 6084, 6095, 6099, 6111, 6122, 6123, 6132, 6140, 6143, 6173, 6177, 6184, 6186, 6192, 6194, 6208, 6217, 6218, 6221, 6223, 6225, 6228, 6230, 6231, 6234, 6236, 6248, 6250, 6251, 6263, 6264, 6269, 6273, 6276, 6279, 6283, 6286, 6291, 6294, 6302, 6309, 6325, 6327, 6335, 6337, 6341, 6347, 6353, 6354, 6359, 6370, 6383, 6387, 6391, 6393, 6408, 6409, 6412, 6418, 6429, 6431, 6434, 6454, 6457, 6461, 6463, 6465, 6467, 6480, 6488, 6490, 6494, 6502, 6505, 6513, 6522, 6525, 6530, 6537, 6540, 6563, 6564, 6565, 6568, 6576, 6582, 6587, 6593, 6594, 6595, 6598, 6599, 6601, 6605, 6619, 6620, 6628, 6630, 6631, 6632, 6638, 6656, 6658, 6674, 6677, 6678, 6688, 6698, 6704, 6726, 6729, 6740, 6755, 6757, 6760, 6762, 6768, 6773, 6774, 6775, 6777, 6791, 6794, 6811, 6812, 6814, 6819, 6824, 6829, 6830, 6831, 6835, 6840, 6841, 6845, 6849, 6850, 6852, 6855, 6859, 6870, 6881, 6885, 6901, 6907, 6908, 6914, 6915, 6916, 6917, 6918, 6919, 6920, 6921, 6922, 6923, 6924, 6925, 6926, 6927, 6928, 6929, 6930, 6931, 6932, 6933, 6934, 6935, 6936, 6937, 6938, 6939, 6940, 6941, 6942, 6943, 6944, 6945, 6946, 6947, 6948, 6949, 6950, 6951, 6952, 6953, 6954, 6955, 6956, 6957, 6958, 6959, 6960, 6961, 6962, 6963, 6964, 6965, 6966, 6967, 6968, 6969, 6970, 6971, 6972, 6973, 6974, 6975, 6976, 6977, 6978, 6979, 6980, 6981, 6982, 6983, 6984, 6985, 6986, 6987, 6988, 6989, 6990, 6991, 6992, 6993, 6994, 6995, 6996, 6997, 6998, 6999, 7000, 7001, 7002, 7003, 7004, 7005, 7006, 7007, 7008, 7009, 7010, 7011, 7012, 7013, 7014, 7015, 7016, 7017, 7018, 7019, 7020, 7021, 7022, 7023, 7024, 7025, 7026, 7027, 7028, 7029, 7030, 7031, 7032, 7033, 7034, 7035, 7036, 7037, 7038, 7039, 7040, 7041, 7042, 7043, 7044, 7045, 7046, 7047, 7048, 7049, 7050, 7051, 7052, 7053, 7054, 7055, 7056, 7057, 7058, 7059, 7060, 7061, 7062, 7063, 7064, 7065, 7066, 7067, 7068, 7069, 7070, 7071, 7072, 7073, 7074, 7075, 7076, 7077, 7078, 7079, 7080, 7081, 7082, 7083, 7084, 7085, 7086, 7087, 7088, 7089, 7090, 7091, 7092, 7093, 7094, 7095, 7096, 7097, 7098, 7099, 7100, 7101, 7102, 7103, 7104, 7105, 7106, 7107, 7108, 7109, 7110, 7111, 7112, 7113, 7114, 7115, 7116, 7117, 7118, 7119, 7120, 7121, 7122, 7123, 7124, 7125, 7126, 7127, 7128, 7129, 7130, 7131, 7132, 7133, 7134, 7135, 7136, 7137, 7138, 7139, 7140, 7141, 7142, 7143, 7144, 7145, 7146, 7147, 7148, 7149, 7150, 7151, 7152, 7153, 7154, 7155, 7156, 7157, 7158, 7159, 7160, 7161, 7162, 7163, 7164, 7165, 7166, 7167, 7168, 7169, 7170, 7171, 7172, 7173, 7174, 7175, 7176, 7177, 7178, 7179, 7180, 7181, 7182, 7183, 7184, 7185, 7186, 7187, 7188, 7189, 7190, 7191, 7192, 7193, 7194, 7195, 7196, 7197, 7198, 7199, 7200, 7201, 7202, 7203, 7204, 7205, 7206, 7207, 7208, 7209, 7210, 7211, 7212, 7213, 7214, 7215, 7216, 7217, 7218, 7219, 7220, 7221, 7222, 7223, 7224, 7225, 7226, 7227, 7228, 7229, 7230, 7231, 7232, 7233, 7234, 7235, 7236, 7237, 7238, 7239, 7240, 7241, 7242, 7243, 7244, 7245, 7246, 7247, 7248, 7249, 7250, 7251, 7252, 7253, 7254, 7255, 7256, 7257, 7258, 7259, 7260, 7261, 7262, 7263, 7264, 7265, 7266, 7267, 7268, 7269, 7270, 7271, 7272, 7273, 7274, 7275, 7276, 7277, 7278, 7279, 7280, 7281, 7282, 7283, 7284, 7285, 7286, 7287, 7288, 7289, 7290, 7291, 7292, 7293, 7294, 7295, 7296, 7297, 7298, 7299, 7300, 7301, 7302, 7303, 7304, 7305, 7306, 7307, 7308, 7309, 7310, 7311, 7312, 7313, 7314, 7315, 7316, 7317, 7318, 7319, 7320, 7321, 7322, 7323, 7324, 7325, 7326, 7327, 7328, 7329, 7330, 7331, 7332, 7333, 7334, 7335, 7336, 7337, 7338, 7339, 7340, 7341, 7342, 7343, 7344, 7345, 7346, 7347, 7348, 7349, 7350, 7351, 7352, 7353, 7354, 7355, 7356, 7357, 7358, 7359, 7360, 7361, 7362, 7363, 7364, 7365, 7366, 7367, 7368, 7369, 7370, 7371, 7372, 7373, 7374, 7375, 7376, 7377, 7378, 7379, 7380, 7381, 7382, 7383, 7384, 7385, 7386, 7387, 7388, 7389, 7390, 7391, 7392, 7393, 7394, 7395, 7396, 7397, 7398, 7399, 7400, 7401, 7402, 7403, 7404, 7405, 7406, 7407, 7408, 7409, 7410, 7411, 7412, 7413, 7414, 7415, 7416, 7417, 7418, 7419, 7420, 7421, 7422, 7423, 7424, 7425, 7426, 7427, 7428, 7429, 7430, 7431, 7432, 7433, 7434, 7435, 7436, 7437, 7438, 7439, 7440, 7441, 7442, 7443, 7444, 7445, 7446, 7447, 7448, 7449, 7450, 7451, 7452, 7453, 7454, 7455, 7456, 7457, 7458, 7459, 7460, 7461, 7462, 7463, 7464, 7465, 7466, 7467, 7468, 7469, 7470, 7471, 7472, 7473, 7474, 7475, 7476, 7477, 7478, 7479, 7480, 7481, 7482, 7483, 7484, 7485, 7486, 7487, 7488, 7489, 7490, 7491, 7492, 7493, 7494, 7495, 7496, 7497, 7498, 7499, 7500, 7501, 7502, 7503, 7504, 7505, 7506, 7507, 7508, 7509, 7510, 7511, 7512, 7513, 7514, 7515, 7516, 7517, 7518, 7519, 7520, 7521, 7522, 7523, 7524, 7525, 7526, 7527, 7528, 7529, 7530, 7531, 7532, 7533, 7534, 7535, 7536, 7537, 7538, 7539, 7540, 7541, 7542, 7543, 7544, 7545, 7546, 7547, 7548, 7549, 7550, 7551, 7552, 7553, 7554, 7555, 7556, 7557, 7558, 7559, 7560, 7561, 7562, 7563, 7564, 7565, 7566, 7567, 7568, 7569, 7570, 7571, 7572, 7573, 7574, 7575, 7576, 7577, 7578, 7579, 7580, 7581, 7582, 7583, 7584, 7585, 7586, 7587, 7588, 7589, 7590, 7591, 7592, 7593, 7594, 7595, 7596, 7597, 7598, 7599, 7600, 7601, 7602, 7603, 7604, 7605, 7606, 7607, 7608, 7609, 7610, 7611, 7612, 7613, 7614, 7615, 7616, 7617, 7618, 7619, 7620, 7621, 7622, 7623, 7624, 7625, 7626, 7627, 7628, 7629, 7630, 7631, 7632, 7633, 7634, 7635, 7636, 7637, 7638, 7639, 7640, 7641, 7642, 7643, 7644, 7645, 7646, 7647, 7648, 7649, 7650, 7651, 7652, 7653, 7654, 7655, 7656, 7657, 7658, 7659, 7660, 7661, 7662, 7663, 7664, 7665, 7666, 7667, 7668, 7669, 7670, 7671, 7672, 7673, 7674, 7675, 7676, 7677, 7678, 7679, 7680, 7681, 7682, 7683, 7684, 7685, 7686, 7687, 7688, 7689, 7690, 7691, 7692, 7693, 7694, 7695, 7696, 7697, 7698, 7699, 7700, 7701, 7702, 7703, 7704, 7705, 7706, 7707, 7708, 7709, 7710, 7711, 7712, 7713, 7714, 7715, 7716, 7717, 7718, 7719, 7720, 7721, 7722, 7723, 7724, 7725,

916, 6920, 6923, 6936, 6937, 6938, 6949, 6964, 6995, 6999, 7005, 7014, 7020, 7021, 7027, 7035, 7036, 7038, 7073, 7076, 7081, 7092, 7098, 7103, 7108, 7113, 7114, 7117, 7146, 7147, 7154, 7162, 7184, 7185, 7201, 7209, 7210, 7216, 7218, 7220, 7228, 7230, 7232, 7241, 7244, 7249, 7251, 7253, 7287, 7290, 7291, 7293, 7294, 7302, 7303, 7308, 7310, 7320, 7329, 7337, 7338, 7344, 7345, 7349, 7356, 7357, 7369, 7371, 7379, 7381, 7383, 7389, 7396, 7402, 7407, 7425, 7430, 7431, 7432, 7440, 7446, 7454, 7457, 7479, 7487, 7490, 7495, 7500, 7503, 7504, 7507, 7508, 7520, 7530, 7535, 7548, 7562, 7567, 7573, 7575, 7583, 7586, 7587, 7590, 7599, 7604, 7610, 7619, 7631, 7656, 7687, 7694, 7698, 7701, 7702, 7704, 7705, 7709, 7711, 7715, 7720, 7723, 7729, 7731, 7735, 7736, 7760, 7770, 7784, 7791, 7801, 7804, 7824, 7831, 7834, 7839, 7850, 7856, 7880, 7900, 7907, 7911, 7922, 7931, 7944, 7948, 7961, 7962, 7972, 7973, 7975, 7976, 7985, 8000, 8015, 8024, 8028, 8029, 8031, 8032, 8050, 8053, 8083, 8098, 8101, 8106, 8131, 8135, 8161, 8162, 8176, 8178, 8186, 8187, 8192, 8198, 8224, 8234, 8249, 8254, 8256, 8258, 8259, 8277, 8293, 8299, 8303, 8304, 8334, 8336, 8340, 8343, 8344, 8347, 8348, 8349, 8352, 8361, 8374, 8378, 8400, 8402, 8405, 8406, 8420, 8425, 8427, 8429, 8433, 8438, 8443, 8448, 8459, 8464, 8470, 8474, 8482, 8486, 8488, 8489, 8495, 8496, 8497, 8499, 8508, 8512, 8530, 8532, 8534, 8536, 8537, 8542, 8544, 8548, 8560, 8569, 8574, 8582, 8583, 8593, 8594, 8601, 8608, 8610, 8622, 8628, 8651, 8662, 8664, 8667, 8670, 8671, 8672, 8677, 8685, 8687, 8688, 8689, 8725, 8749, 8759, 8766, 8770, 8773, 8775, 8782, 8792, 8794, 8796, 8798, 8803, 8807, 8811, 8813, 8820, 8842, 8843, 8844, 8849, 8860, 8867, 8872, 8882, 8883, 8889, 8891, 8896, 8898, 8900, 8934, 8936, 8938, 8957, 8963, 8971, 8974, 8979, 8991, 8993, 8996, 8997, 9008, 9009, 9016, 9019, 9021, 9025, 9028, 9035, 9037, 9042, 9046, 9050, 9066, 9068, 9072, 9073, 9077, 9093, 9094, 9098, 9099, 9101, 9102, 9105, 9113, 9133, 9134, 9138, 9139, 9144, 9163, 9167, 9168, 9168, 9175, 9176, 9177, 9179, 9180, 9182, 9183, 9189, 9197, 9198, 9203, 9207, 9229, 9231, 9242, 9248, 9261, 9266, 9273, 9284, 9286, 9287, 9288, 9291, 9292, 9293, 9296, 9310, 9318, 9327, 9353, 9356, 9364, 9370, 9371, 9374, 9376, 9378, 9384, 9387, 9388, 9394, 9397, 9398, 9399, 9404, 9421, 9423, 9434, 9436, 9437, 9438, 9439, 9441, 9444, 9448, 9456, 9458, 9465, 9472, 9474, 9475, 9487, 9508, 9511, 9513, 9514, 9519, 9526, 9527, 9530, 9533, 9535, 9538, 9539, 9540, 9552, 9563, 9564, 9565, 9567, 9568, 9570, 9580, 9581, 9586, 9591, 9600, 9624, 9628, 9634, 9636, 9645, 9646, 9653, 9660, 9665, 9673, 9676, 9687, 9695, 9698, 9705, 9707, 9722, 9727, 9734, 9739, 9743, 9754, 9762, 9765, 9780, 9784, 9803, 9819, 9821, 9851, 9860, 9871, 9872, 9899, 9910, 9922, 9925, 9927, 9930, 9938, 9968, 9970, 9992, 9998, 10005, 10006, 10008, 10010, 10017, 10056, 10061, 10074, 10083, 10088, 10089, 10093, 10094, 10099, 10107, 10110, 10120, 10123, 10125, 10135, 10136, 10139, 10180, 10196, 10197, 10215, 10221, 10235, 10239, 10242, 10273, 10276, 10289, 10290, 10291, 10293, 10297, 10308, 10313, 10317, 10325, 10329, 10330, 10341, 10350, 10357, 10359, 10360, 10361, 10368, 10371, 10373, 10380, 10398, 10400, 10422, 10424, 10439, 10448, 10452, 10457, 10458, 10460, 10465, 10472, 10474, 10482, 10486, 10487, 10488, 10494, 10495, 10500, 10508, 10514, 10515, 10517, 10525, 10536, 10540, 10557, 10563, 10573, 10576, 10579, 10580, 10591, 10621, 10652, 10654, 10655, 10682, 10687, 10689, 10708, 10721, 10750, 10800, 10815, 10827, 10836, 10838, 10840, 10853, 10863, 10996, 10997, 11035, 11050, 11075, 11080, 11088, 11097, 11106, 11132, 11137, 11141, 11154, 11157, 11159, 11161, 11171, 11177, 11187, 11194, 11205, 11212, 11216, 11223, 11244, 11257, 11271, 11302, 11309, 11350, 11359, 11361, 11362, 11376, 11381, 11394, 11395, 11399, 11405, 11409, 11412, 11413, 11422, 11424, 11454, 11459, 11472, 11478, 11483, 11484, 11486, 11557, 11559, 11561, 11569, 11593, 11599, 11605, 11606, 11620, 11625, 11643, 11660, 11677, 11686, 11708, 11741, 11743, 11745, 11748, 11770, 11790, 11801, 11810, 11813, 11816, 11822, 11830, 11835, 11852, 11862, 11905, 11906, 11910, 11916, 11923, 11925, 11926, 11928, 11958, 11959, 11968, 12011, 12035, 12048, 12050, 12054, 12087, 12088, 12122, 12152, 12158, 12161, 12164, 12170]

(35000, 1886) (15000, 1886)

In [0]:

```
#https://stackabuse.com/cross-validation-and-grid-search-for-model-selection-in-python/
#https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

#Initialising Classifier
classifier = linear_model.SGDClassifier(loss='log', n_jobs=-1, class_weight='balanced')

#Brute force approach for finding best K value
parameters = {'alpha':[0.0001,0.001,0.01,0.1,1,10,100,1000,10000]}

#Training the model on train data
SGD_set5 = GridSearchCV(classifier, parameters, return_train_score=True, cv=3, scoring='roc_auc', n_jobs=-1)
SGD_set5.fit(df_train_set5, y_train_set5)
```

Out[0]:

```
GridSearchCV(cv=3, error_score='raise-deprecating',
             estimator=SGDClassifier(alpha=0.0001, average=False,
                                     class_weight='balanced',
                                     early_stopping=False, epsilon=0.1,
                                     eta0=0.0, fit_intercept=True,
                                     l1_ratio=0.15, learning_rate='optimal',
                                     loss='log', max_iter=1000,
                                     n_iter_no_change=5, n_jobs=-1,
                                     penalty='l2', power_t=0.5,
                                     random_state=None, shuffle=True, tol=0.001,
                                     validation_fraction=0.1, verbose=0,
                                     warm_start=False)
```

```

warm_start=False,
iid='warn', n_jobs=-1,
param_grid={'alpha': [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000,
                      10000]},
pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
scoring='roc_auc', verbose=0)

```

In [0]:

```

#https://matplotlib.org/api/\_as\_gen/matplotlib.pyplot.plot.html
print(SGD_set5.best_params_) #Gives the best value of K from the given neighbor range
print(parameters['alpha'],SGD_set5.cv_results_['mean_train_score'],
SGD_set5.cv_results_['mean_test_score'])

```

```

log_params = []
for i in parameters['alpha']:
    log_params.append(math.log(i))

```

```
print(log_params)
```

```

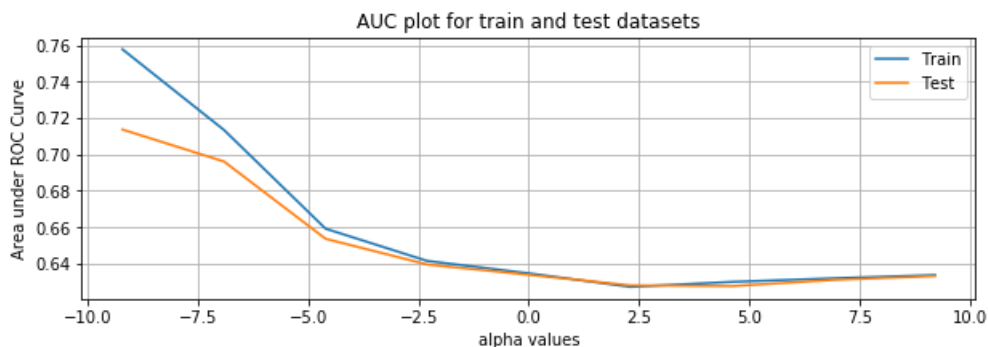
plt.figure(figsize=(10,3))
plt.plot(log_params,SGD_set5.cv_results_['mean_train_score'], label="Train")
plt.plot(log_params,SGD_set5.cv_results_['mean_test_score'], label="Test")
plt.title('AUC plot for train and test datasets')
plt.xlabel('alpha values')
plt.ylabel('Area under ROC Curve')
plt.legend()
plt.grid()
plt.show()
plt.close()

```

```

{'alpha': 0.0001}
[0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000] [0.75772612 0.71335907 0.65908372 0.64136165 0.
.63455662 0.62708241
0.62983176 0.63181806 0.63359657] [0.71361689 0.69603882 0.65353856 0.63941107 0.63371711
0.62793358
0.62751635 0.63091356 0.6330499 ]
[-9.210340371976182, -6.907755278982137, -4.605170185988091, -2.3025850929940455, 0.0,
2.302585092994046, 4.605170185988092, 6.907755278982137, 9.210340371976184]

```



In [0]:

```

#https://datascience.stackexchange.com/questions/22762/understanding-predict-proba-from-multioutputclassifier
#https://stackoverflow.com/questions/34894587/should-we-plot-the-roc-curve-for-each-class
#https://stackoverflow.com/questions/20459536/convert-pandas-dataframe-to-sparse-numpy-matrix-directly

from sklearn.metrics import roc_curve, auc
from scipy.sparse import csr_matrix

#training the model on the best K value found in the above result
final_SGD_set5 = linear_model.SGDClassifier(loss='log', alpha=0.001, n_jobs=-1, class_weight='balanced')
final_SGD_set5.fit(df_train_set5,y_train_set5)

final_df_train_set5=csr_matrix(df_train_set5.values)
final_df_test_set5=csr_matrix(df_test_set5.values)

y_train_set5_pred=[]

```

```

y_test_set5_pred=[]

#ROC curve function takes the actual values and the predicted probabilities of the positive class
for i in range(0,final_df_train_set5.shape[0]):
    y_train_set5_pred.extend(final_SGD_set5.predict_proba(final_df_train_set5[i][:,1]) #[:,1] gives the probability for class 1

for i in range(0,final_df_test_set5.shape[0]):
    y_test_set5_pred.extend(final_SGD_set5.predict_proba(final_df_test_set5[i][:,1])

```

In [0]:

```

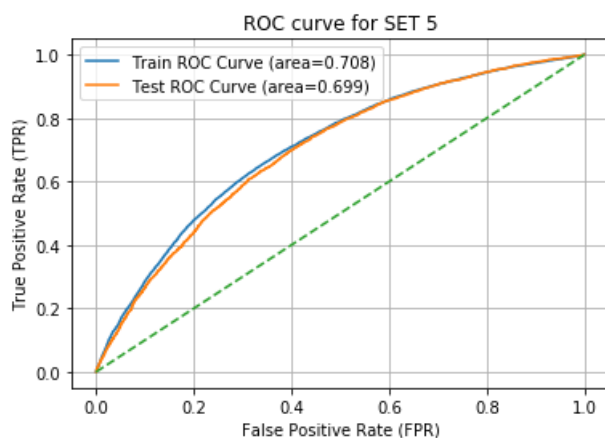
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#Calculating FPR and TPR for train and test data
train_set5_fpr, train_set5_tpr, train_set5_thresholds = roc_curve(y_train_set5, y_train_set5_pred)
test_set5_fpr, test_set5_tpr, test_set5_thresholds = roc_curve(y_test_set5, y_test_set5_pred)

#Calculating AUC for train and test curves
roc_auc_set5_train=auc(train_set5_fpr,train_set5_tpr)
roc_auc_set5_test=auc(test_set5_fpr,test_set5_tpr)

plt.plot(train_set5_fpr, train_set5_tpr, label="Train ROC Curve (area=%0.3f)" % roc_auc_set5_train)
plt.plot(test_set5_fpr, test_set5_tpr, label="Test ROC Curve (area=%0.3f)" % roc_auc_set5_test)
plt.plot([0,1],[0,1],linestyle='--')
plt.legend()
plt.xlabel("False Positive Rate (FPR)")
plt.ylabel("True Positive Rate (TPR)")
plt.title("ROC curve for SET 5")
plt.grid()
plt.show()
plt.close()

```



In [0]:

```

#https://stackoverflow.com/questions/35572000/how-can-i-plot-a-confusion-matrix
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
#https://datatofish.com/confusion-matrix-python/

from sklearn.metrics import confusion_matrix as cf_mx

expected_set5_train = y_train_set5.values
predicted_set5_train = final_SGD_set5.predict(final_df_train_set5)

expected_set5_test = y_test_set5.values
predicted_set5_test = final_SGD_set5.predict(final_df_test_set5)

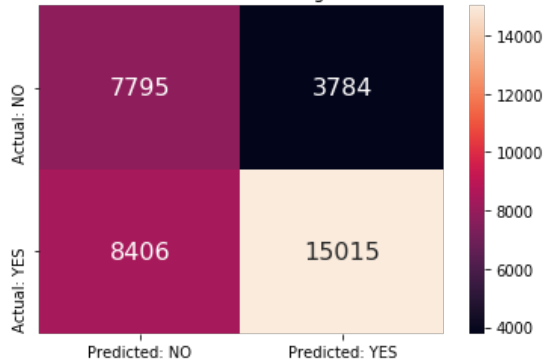
plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_set5_train, predicted_set5_train)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train, annot=True,annot_kws={"size": 16}, fmt='g')

```

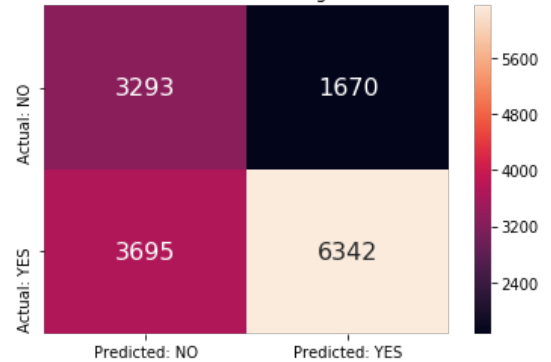
```
plt.title('Confusion matrix for train data using TFIDF W2V(SET 5)')

plt.subplot(1,2,2)
cmdf_test=cf_mx(expected_set5_test, predicted_set5_test)
df_cm_test = pd.DataFrame(cmdf_test, range(2),range(2))
df_cm_test.columns = ['Predicted: NO','Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True,annot_kws={"size": 16}, fmt='g')
plt.title('Confusion matrix for test data using TFIDF W2V(SET 5)')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()
```

Confusion matrix for train data using TFIDF W2V(SET 5)



Confusion matrix for test data using TFIDF W2V(SET 5)



#### Inference:

- Using only top 1960 features as the rest of the features are of zero importance
- The model seems to be produced better results with better TPR and TNR(principal diagonal elements)

## 3. Conclusions

In [9]:

```
#http://zetcode.com/python/prettytable/
```

```
from prettytable import PrettyTable
```

```
x = PrettyTable()
```

```
x.field_names = ["Vectorizer", "Model", "max_depth(DT)/alpha(SVM)", "min_samples_split", "AUC(Train)", "AUC(Test)"]
```

```
x.add_row(["BoW", "Brute", 10, 500, 0.731, 0.698])
```

```
x.add_row(["TFIDF", "Brute", 10, 500, 0.737, 0.689])
```

```
x.add_row(["W2V", "Brute", 10, 500, 0.733, 0.680])
```

```
x.add_row(["TFIDF AVG W2V", "Brute", 10, 500, 0.735, 0.680])
```

```
x.add_row(["TFIDF Linear SVM \n(Top features)", "Brute", 0.0001, 'NA', 0.708, 0.699])
```

```
print(x)
```

Vectorizer	Model	max_depth(DT)/alpha(SVM)	min_samples_split	AUC(Train)	AUC(Test)
BoW	Brute	10	500	0.731	0.698
TFIDF	Brute	10	500	0.737	0.689
W2V	Brute	10	500	0.733	0.68
TFIDF AVG W2V	Brute	10	500	0.735	0.68
TFIDF Linear SVM (Top features)	Brute	0.0001	NA	0.708	0.699

- Most of the projects which were falsely identified as positive, were posted by teachers with very less or zero number of previously posted projects

previously posted projects.

- SET 1 using Bag of Words has produced better results both on train and test data.

In [0]: