

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. Example: p036502
<code>project_title</code>	<ul style="list-style-type: none">••	Title of the project. Examples: <code>Art Will Make You Happy!</code> <code>First Grade Fun</code>
<code>project_grade_category</code>	<ul style="list-style-type: none">••••	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">•••••••••	One or more (comma-separated) subject categories for the project from the following enumerated list of values: <code>Applied Learning</code> <code>Care & Hunger</code> <code>Health & Sports</code> <code>History & Civics</code> <code>Literacy & Language</code> <code>Math & Science</code> <code>Music & The Arts</code> <code>Special Needs</code> <code>Warmth</code> Examples: <ul style="list-style-type: none">• <code>Music & The Arts</code>• <code>Literacy & Language, Math & Science</code>
<code>school_state</code>		State where school is located (Two-letter U.S. postal code). Example: WY
<code>project_subject_subcategories</code>	<ul style="list-style-type: none">••	One or more (comma-separated) subject subcategories for the project. Examples: <code>Literacy</code> <code>Literature & Writing, Social Sciences</code>
<code>project_resource_summary</code>	<ul style="list-style-type: none">•	An explanation of the resources needed for the project. Example: <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. Example: 2016-04-28 12:43:56.245
teacher_id	A unique identifier for the teacher of the proposed project. Example: bdf8baa8fedef6bfeec7ae4ff1c15c56
teacher_prefix	Teacher's title. One of the following enumerated values: <ul style="list-style-type: none"> nan Dr. Mr. Mrs. Ms. Teacher.
teacher_number_of_previously_posted_projects	Number of project applications previously submitted by the same teacher. Example: 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. Example: p036502
description	Description of the resource. Example: Tenor Saxophone Reeds, Box of 25
quantity	Quantity of the resource required. Example: 3
price	Price of the resource required. Example: 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 [2]:

```
%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 [3]:

```

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%3Aawg%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 [4]:

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

'06 Implement SGD.ipynb'	Db-IMDB.db
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'
9_DonorsChoose_RF_final.ipynb	train_data.csv

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 [6]:

```
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 [7]:

```
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 [8]:

```
# 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[8]:

	Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_is_approved
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 [9]:

```
print("Number of data points in train data", resource_data.shape)
print(resource_data.columns.values)
resource_data.head(2)
```

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

Out[9]:

	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" => "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 placeing 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 [13]:

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

Out[13]:

Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_
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 [14]:

```

# 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 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

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 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 [16]:

```
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 do not have a lot of academic practice outside of the school day. They love coming to school every day and are eager to learn. They work very hard and are so excited when they master new concepts.

```
\r\n At my school site we strive to make the most of every minute during the school day in order
to ensure students are able to learn and feel successful. We know that the time we have with them
is very precious! I am asking for the mini white boards and reusable write and wipe pockets in orde
r to help me monitor my students thinking and learning. Often times, when work is done on
worksheets the feedback to students is not meaningful because it can take awhile to give each stud
ent individual feed back. The white boards and write and wipe pockets will give students a way to
show written responses while we are gathered at the carpet together. This will allow me to give im
mediate feedback to students and then can modify their responses right then and there. This will l
ead to more meaningful learning and processing.nannan
```

In [17]:

```
# \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)
```

My school is in a low socio-economic area with a high ELL population. The students in my classroom do not have a lot of academic practice outside of the school day. They love coming to school every day and are eager to learn. They work very hard and are so excited when they master new concepts. At my school site we strive to make the most of every minute during the school day in order to ensure students are able to learn and feel successful. We know that the time we have with them is very precious! I am asking for the mini white boards and reusable write and wipe pockets in order to help me monitor my students thinking and learning. Often times, when work is done on worksheets the feedback to students is not meaningful because it can take awhile to give each student individual feedback. The white boards and write and wipe pockets will give students a way to show written responses while we are gathered at the carpet together. This will allow me to give immediate feedback to students and then can modify their responses right then and there. This will lead to more meaningful learning and processing.nannan

In [18]:

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

My school is in a low socio economic area with a high ELL population The students in my classroom do not have a lot of academic practice outside of the school day They love coming to school every day and are eager to learn They work very hard and are so excited when they master new concepts At my school site we strive to make the most of every minute during the school day in order to ensure students are able to learn and feel successful We know that the time we have with them is very precious I am asking for the mini white boards and reusable write and wipe pockets in order to help me monitor my students thinking and learning Often times when work is done on worksheets the feedback to students is not meaningful because it can take awhile to give each student individual feedback The white boards and write and wipe pockets will give students a way to show written responses while we are gathered at the carpet together This will allow me to give immediate feedback to students and then can modify their responses right then and there This will lead to more meaningful learning and processing nannan

In [0]:

```
# https://gist.github.com/sebleier/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',\
            'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'e
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 [20]:

```

# Combining all the above students
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('\n', ' ')
    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 [00:25<00:00, 1936.67it/s]

In [21]:

```

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

# after preprocessing
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[21]:

'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 [22]:

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

Out[22]:

Unnamed: 0	id	teacher_id	teacher_prefix	school_state	Date	project_grade_category	project_title
0							

In [23]:

```
#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 [24]:

```
# 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, 4256.38it/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 [27]:

```
#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[27]:

```
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 [28]:

```
#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 [29]:

```
project_data.columns
```

Out[29]:

```
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 [30]:

```
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 [31]:

```
# 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 [32]:

```
# 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 [33]:

```
# 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 [34]:

```
# 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 [35]:

```
# 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 [36]:

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

# Now standardize the data with above mean and variance.
price_train_standardized = price_scalar.transform(df_train['price'].values.reshape(-1, 1))
price_test_standardized = price_scalar.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 : 312.3405677142857, Standard deviation : 382.99294145436824
101
```

2.2.2.2 Vectorizing no. of previously posted projects

In [37]:

```
# 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_scalar = StandardScaler()
prev_proj_scalar.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_scalar.mean_[0]}, Standard deviation : {np.sqrt(prev_proj_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
prev_proj_train_standardized = prev_proj_scalar.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 : 10.4772, Standard deviation : 26.846264866883384
102

2.2.2.3 Vectorizing word counts of project title

In [38]:

```
# 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 : 3.6676285714285712, Standard deviation : 1.5420257842149312
103

2.2.2.4 Vectorizing word counts of essay text

In [39]:

```
# 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 : 136.65925714285714, Standard deviation : 35.69315352544547
104

2.2.2.5 Vectorizing sentimental scores of project essays

In [40]:


```
# 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 : 0.9582661000000001, Standard deviation : 0.15392388114423403
105

2.2.2.6 Vectorizing Quantity

In [41]:

```
# 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 : 17.70797142857143, Standard deviation : 27.69658368616783
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 [42]:

```
# We are considering only the words which appeared in at least 10 documents(rows or projects).
vectorizer = CountVectorizer(min_df=10)
```

```

vectorizer = CountVecorizer(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, 10501) (15000, 10501)
10607

In [43]:

```

# 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, 1584) (15000, 1584)
12191

2.3.1.2 TFIDF vectorizer for essay text

In [44]:

```

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, 10501) (15000, 10501)
10607

In [45]:

```

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 encodig ",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 encodig (35000, 1584) (15000, 1584)
12191

2.3.1.3 Using Pretrained models: Avg W2V vectorizer

In [46]:

```

'''def loadGloveModel(gloveFile):

```

```
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[46]:

```
'def loadGloveModel(gloveFile):\n    print ("Loading Glove Model")\n    f = open(gloveFile,\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 = loadGloveModel(\n'drive/My Drive/Colab Notebooks/glove.6B.50d.txt')'
```

In [47]:

```
'''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[47]:

```
'words = []\nfor i in preprocessed_essays:\n    words.extend(i.split(\' \''))\n\nfor i in preprocessed_title:\n    words.extend(i.split(\' \''))\n\nprint("all the words in the coupus", len(words))\n\nwords = set(words)\n\nprint("the unique words in the coupus", len(words))\n\ninter_words = set(model.keys()).intersection(words)\n\nprint("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 = {}\nwords_glove = set(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# stronging variables into pickle files python : http://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-python/\n\nimport pickle\n\nwith open(\n'drive/My Drive/Colab Notebooks/glove_vectors_50', \n'wb') as f:\n    pickle.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)
    glove_words = set(model.keys())
```

In [49]:

```
# 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 [00:07<00:00, 4980.77it/s]

35000
50

In [50]:

```
# 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 [00:03<00:00, 4873.04it/s]

15000
50

In [51]:

```
# 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]
            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, 101098.26it/s]

35000
50

In [52]:

```

# 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, 101492.12it/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 [54]:

```

# 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 += (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 [01:17<00:00, 450.30it/s]

35000
50

In [55]:

```

# 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 [00:36<00:00, 410.56it/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 [57]:

```

# 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)

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

```

100%|██████████| 35000/35000 [00:01<00:00, 29391.85it/s]

35000
50

In [58]:

```

# 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, 29752.44it/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 [59]:

```

#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_essay_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 'scipy.sparse.csr.csr_matrix'> <class 'scipy.sparse.csr.csr_matrix'>
(35000, 12191) <class 'scipy.sparse.coo.coo_matrix'> (35000,) <class 'pandas.core.series.Series'>
(15000, 12191) <class 'scipy.sparse.coo.coo_matrix'> (15000,) <class 'pandas.core.series.Series'>

```

In [103]:

```

#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],
              '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[103]:

```

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],
                         'min_samples_split': [5, 10, 100, 500]},
             pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
             scoring='roc_auc', verbose=0)

```


In [118]:

```
#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

print(params_train)
print(params_test)

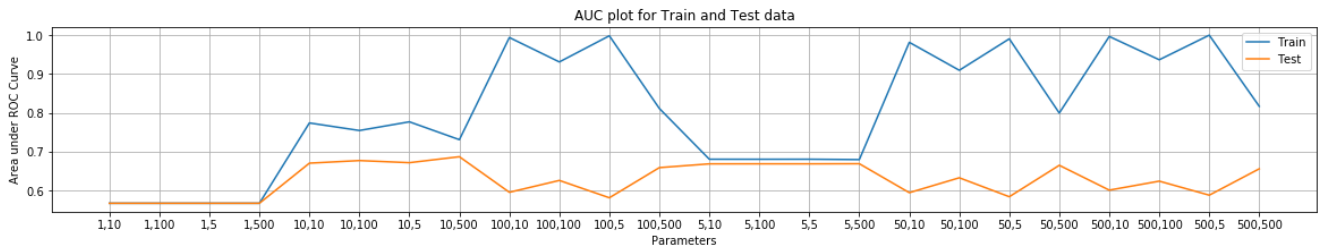
params_train = {k: params_train[k] for k in sorted(params_train)}
params_test = {k: params_test[k] for k in sorted(params_test)}

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 Train and Test data')
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}]
{'1,5': 0.5670500292815107, '1,10': 0.5670500292815107, '1,100': 0.5670500292815107, '1,500': 0.56
70500292815107, '5,5': 0.6804618726507476, '5,10': 0.6804429118966453, '5,100': 0.680382215402577,
'5,500': 0.6794727464576211, '10,5': 0.7767445911729507, '10,10': 0.7738428789322599, '10,100': 0.
7543854128411694, '10,500': 0.7307890246865005, '50,5': 0.9903785328072484, '50,10':
0.9813619723129179, '50,100': 0.9094075164326737, '50,500': 0.7993060562134605, '100,5':
0.9981935314852495, '100,10': 0.9937539214628736, '100,100': 0.930883981620213, '100,500':
0.81111634574099, '500,5': 0.9997988154108247, '500,10': 0.996489963714882, '500,100':
0.9364665972209654, '500,500': 0.8166221823481014}
{'1,5': 0.5670498591278498, '1,10': 0.5670498591278498, '1,100': 0.5670498591278498, '1,500': 0.56
70498591278498, '5,5': 0.6685847498297001, '5,10': 0.6685421065904512, '5,100':
0.6686370537289972, '5,500': 0.6688996946436353, '10,5': 0.6715387176509116, '10,10':
0.6703082966128884, '10,100': 0.6770182563369168, '10,500': 0.6868834465478598, '50,5':
0.5839130482821885, '50,10': 0.5943504150948153, '50,100': 0.6326879721566375, '50,500':
0.66484448069417904, '100,5': 0.5812220168849842, '100,10': 0.5953569269639721, '100,100':
0.6258774401753013, '100,500': 0.6588262415132163, '500,5': 0.5879131460282525, '500,10':
0.6008848164011591, '500,100': 0.6240177067012266, '500,500': 0.6557915176464189}
{'1,10': 0.5670500292815107, '1,100': 0.5670500292815107, '1,5': 0.5670500292815107, '1,500': 0.56
70500292815107, '10,10': 0.7738428789322599, '10,100': 0.7543854128411694, '10,5':
0.7767445911729507, '10,500': 0.7307890246865005, '100,10': 0.9937539214628736, '100,100':
0.930883981620213, '100,5': 0.9981935314852495, '100,500': 0.81111634574099, '5,10':
0.6804429118966453, '5,100': 0.680382215402577, '5,5': 0.6804618726507476, '5,500':
0.6794727464576211, '50,10': 0.9813619723129179, '50,100': 0.9094075164326737, '50,5':
0.9903785328072484, '50,500': 0.7993060562134605, '500,10': 0.996489963714882, '500,100':
```

```
0.9364665972209654, '500,5': 0.9997988154108247, '500,500': 0.8166221823481014}
{'1,10': 0.5670498591278498, '1,100': 0.5670498591278498, '1,5': 0.5670498591278498, '1,500': 0.5670498591278498, '10,10': 0.6703082966128884, '10,100': 0.6770182563369168, '10,5': 0.6715387176509116, '10,500': 0.6868834465478598, '100,10': 0.5953569269639721, '100,100': 0.6258774401753013, '100,5': 0.5812220168849842, '100,500': 0.6588262415132163, '5,10': 0.6685421065904512, '5,100': 0.6686370537289972, '5,5': 0.6685847498297001, '5,500': 0.6688996946436353, '50,10': 0.5943504150948153, '50,100': 0.6326879721566375, '50,5': 0.5839130482821885, '50,500': 0.6648448069417904, '500,10': 0.6008848164011591, '500,100': 0.6240177067012266, '500,5': 0.5879131460282525, '500,500': 0.6557915176464189}
```



In [143]:

```
#https://seaborn.pydata.org/generated/seaborn.heatmap.html

#print(DT_BoW.cv_results_['params'])

test_BOW_md_params = []
test_BOW_ms_params = []
test_BOW_results = []

for i in DT_BoW.cv_results_['params']:
    test_BOW_md_params.append(i['max_depth'])
    test_BOW_ms_params.append(i['max_depth'])

for i in DT_BoW.cv_results_['mean_train_score']:
    test_BOW_results.append(i)

print(len(test_BOW_ms_params), len(test_BOW_md_params), len(test_BOW_results))

import numpy as np;
import seaborn as sns;
import pandas as pd

val=[[0.5670500292815107, 0.5670500292815107, 0.5670500292815107, 0.5670500292815107],
[0.6804618726507476, 0.6804429118966453, 0.680382215402577, 0.6794727464576211],
[0.7767445911729507, 0.7738428789322599, 0.7543854128411694, 0.7307890246865005],
[0.9903785328072484, 0.9813619723129179, 0.9094075164326737, 0.7993060562134605],
[0.9981935314852495, 0.9937539214628736, 0.930883981620213, 0.81111634574099],
[0.9997988154108247, 0.996489963714882, 0.9364665972209654, 0.8166221823481014]]

heatmap_df=pd.DataFrame(val)
#print(heatmap_df)
heatmap_df.index=parameters['max_depth']
heatmap_df.columns=parameters['min_samples_split']

plt.subplots(figsize=(20,4))
plt.subplot(1,2,1)
sns.heatmap(heatmap_df, annot=True,annot_kws={"size": 10}, fmt='g')
plt.title('AUC plot for Train data')
plt.subplots_adjust(wspace=0.5)

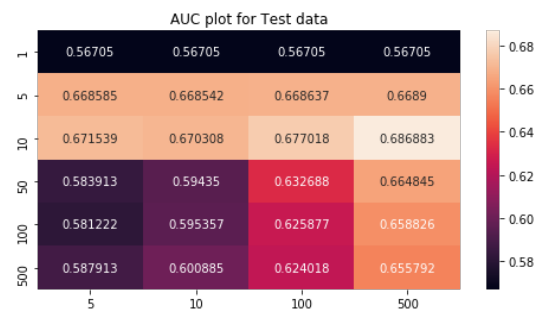
val=[[0.5670498591278498, 0.5670498591278498, 0.5670498591278498, 0.5670498591278498],
[0.6685847498297001, 0.6685421065904512, 0.6686370537289972, 0.6688996946436353],
[0.6715387176509116, 0.6703082966128884, 0.6770182563369168, 0.6868834465478598],
[0.5839130482821885, 0.5943504150948153, 0.6326879721566375, 0.6648448069417904],
[0.5812220168849842, 0.5953569269639721, 0.6258774401753013, 0.6588262415132163],
[0.5879131460282525, 0.6008848164011591, 0.6240177067012266, 0.6557915176464189]]

heatmap_df=pd.DataFrame(val)
#print(heatmap_df)
heatmap_df.index=parameters['max_depth']
heatmap_df.columns=parameters['min_samples_split']

plt.subplot(1,2,2)
sns.heatmap(heatmap_df, annot=True,annot_kws={"size": 10}, fmt='g')
```

```
plt.title('AUC plot for Test data')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()
```

24 24 24



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 probability 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]:

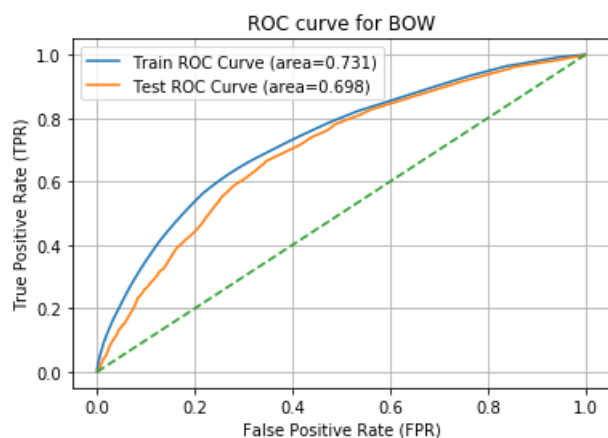
```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#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

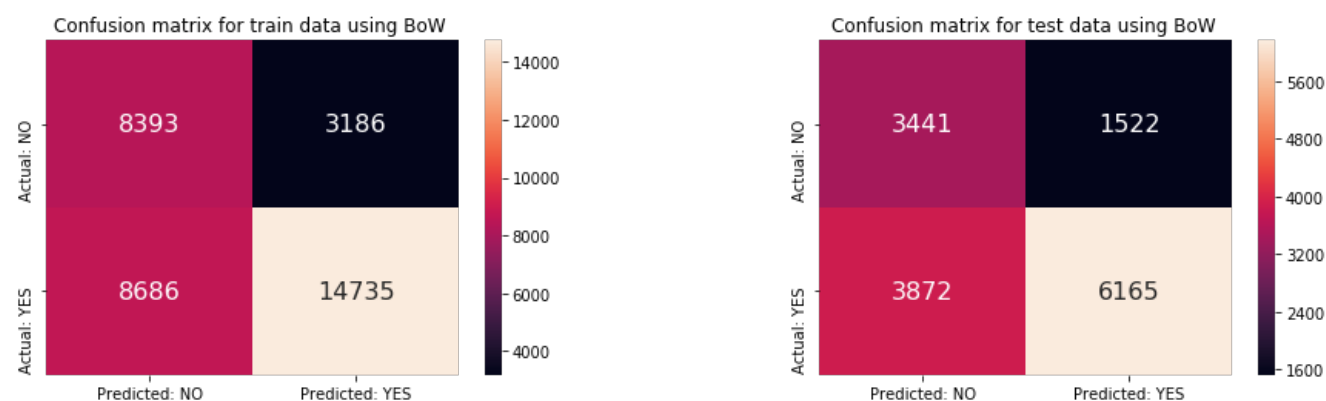
expected_train = y_train.values
predicted_train = final_DT_BoW.predict(x_train)

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\natives-identified-by-a-conf\n#https://stackoverflow.com/questions/31324218/scikit-learn-how-to-obtain-true-positive-true-negati\nve-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, 318
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, 379
7, 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, 5998, 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, 7599, 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, 8683, 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]

```
#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)
```

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)
```

In [0]:

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

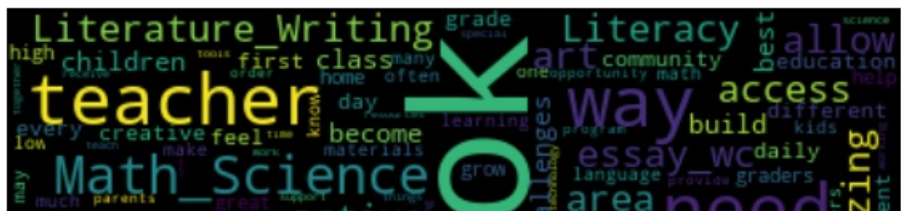
fp_features = []

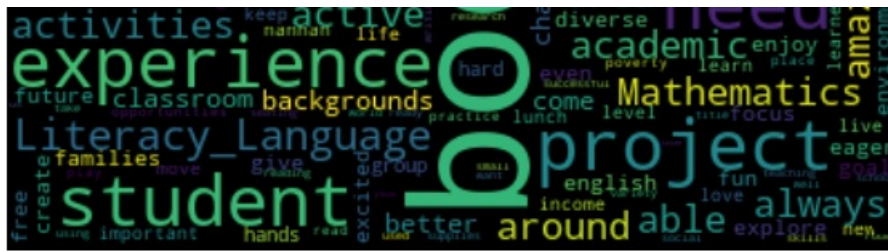
print(x_test_df.shape[0])
for i in range(x_test_df.shape[0]):
    if x_test_df[i] > 250:
        fp_features.append(feature_names_bow[i])

print(len(fp_features))
```

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()
```





Box plot for Price feature of False positives

In [0]:

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

1522

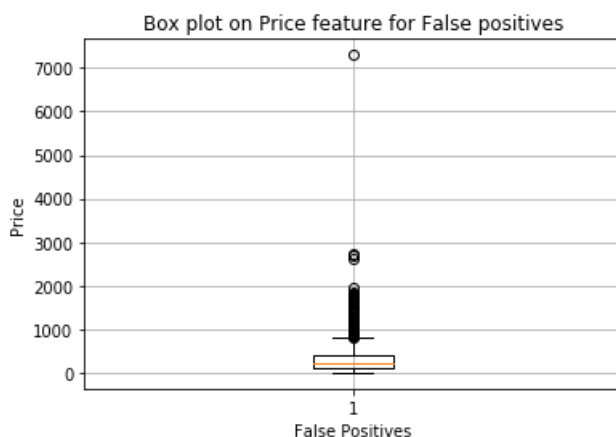
```
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'Date', 'project_grade_category', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects',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

In [0]:

```
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'> (1522,)
```



- **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(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')
```

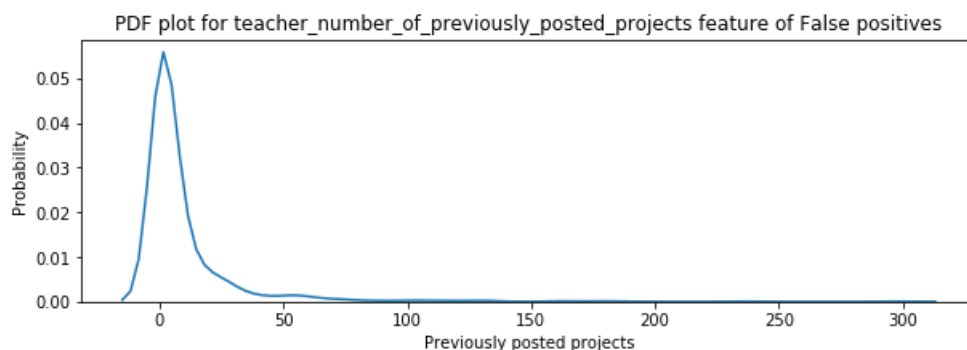
In [0]:

```
#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,
                    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,j 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)

```

```

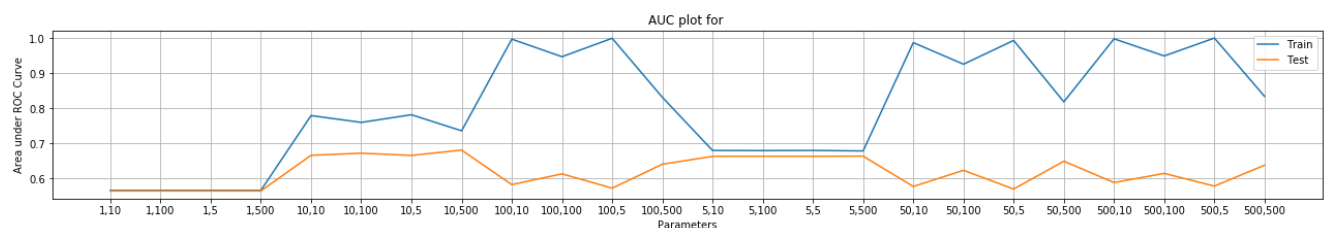
plt.figure(figsize=(22,3))
plt.plot(params_train.keys(),params_train.values(), label="Train")
plt.plot(params_test.keys(),params_test.values(), label="Test")
plt.title('AUC plot for ')
plt.xlabel('Parameters')
plt.ylabel('Area under ROC Curve')
plt.legend()
plt.grid()
plt.show()
plt.close()

```

```

{'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': 0.5658447630496513, '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': 0.5641422578751661, '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 [144]:

```

import numpy as np;
import seaborn as sns;
import pandas as pd

```

```

val=[[0.5658447630496513, 0.5658447630496513, 0.5658447630496513, 0.5658447630496513],
[0.6796173471044051, 0.679598389536984, 0.6791455017824047, 0.6781862048164735],
[0.7813826150347807, 0.7790041149886924, 0.7595613115623332, 0.7355770301342763],
[0.9931276554327202, 0.9869151939333026, 0.9254593650937502, 0.8181435325685966],
[0.9993509109059684, 0.9968314300989171, 0.946771677437496, 0.8311269209917208],
[0.999905702924585, 0.9978182294009603, 0.9489810922741944, 0.833743704156233]]

```

```

heatmap_df=pd.DataFrame(val)
#print(heatmap_df)
heatmap_df.index=parameters['max_depth']

```

```

heatmap_df.columns=parameters['min_samples_split']

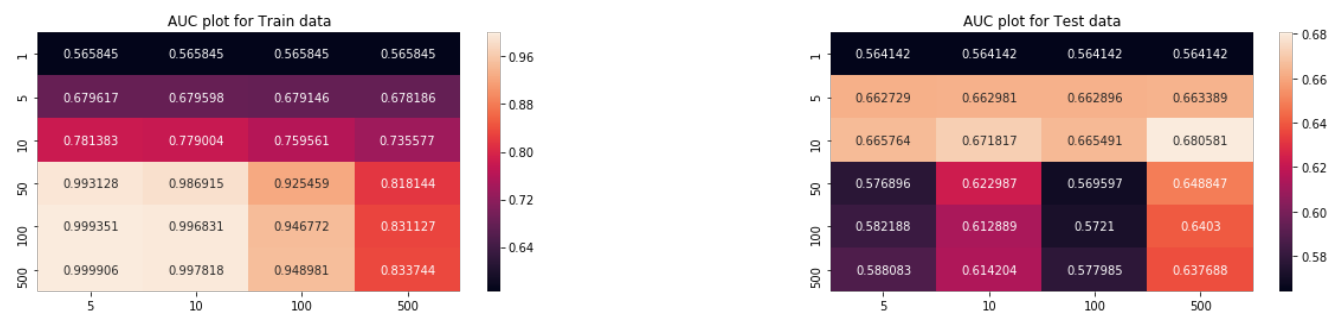
plt.subplots(figsize=(20,4))
plt.subplot(1,2,1)
sns.heatmap(heatmap_df, annot=True,annot_kws={"size": 10}, fmt='g')
plt.title('AUC plot for Train data')
plt.subplots_adjust(wspace=0.5)

val=[[0.5641422578751661, 0.5641422578751661, 0.5641422578751661, 0.5641422578751661],
[0.6627287854950775, 0.6629807427116557, 0.6628963663197197, 0.6633888496465025],
[0.6657641903224826, 0.6718169254258323, 0.6654909854554135, 0.6805807622379173],
[0.5768960913443291, 0.6229874426809158, 0.5695971270088239, 0.6488472378557606],
[0.5821878962403411, 0.6128886302413883, 0.5720998585817175, 0.6403000911203013],
[0.5880832937497642, 0.6142040545290168, 0.5779854299978125, 0.637688299731681]]

heatmap_df=pd.DataFrame(val)
#print(heatmap_df)
heatmap_df.index=parameters['max_depth']
heatmap_df.columns=parameters['min_samples_split']

plt.subplot(1,2,2)
sns.heatmap(heatmap_df, annot=True,annot_kws={"size": 10}, fmt='g')
plt.title('AUC plot for Test data')
plt.subplots_adjust(wspace=0.5)
plt.show()
plt.close()

```



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

```

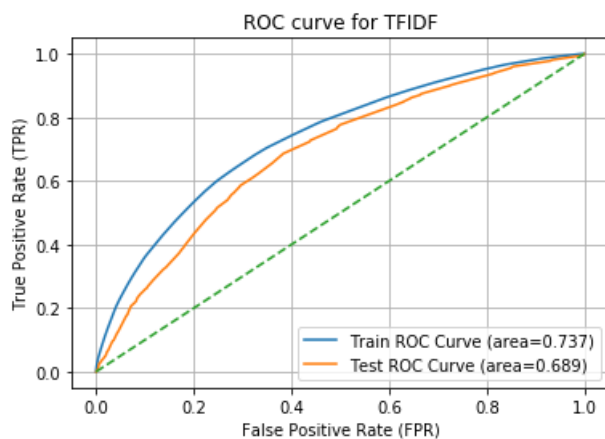
```

#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)

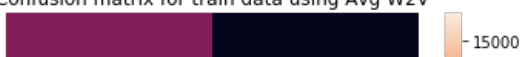
expected_test_tfidf = y_test_tfidf.values
predicted_test_tfidf = final_DT_tfidf.predict(x_test_tfidf)

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')

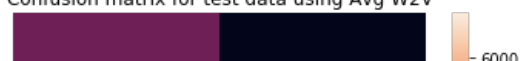
plt.subplot(1,2,2)
cmdf_test=cf_mx(expected_test_tfidf, predicted_test_tfidf)
df_cm_test = pd.DataFrame(cmdf_test, range(2),range(2))
df_cm_test.columns = ['Predicted: NO','Predicted: YES']
df_cm_test = df_cm_test.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_test, annot=True,annot_kws={"size":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

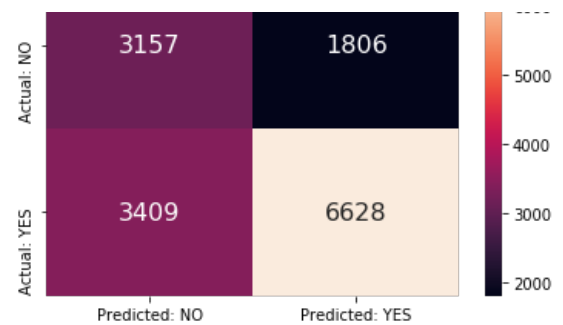
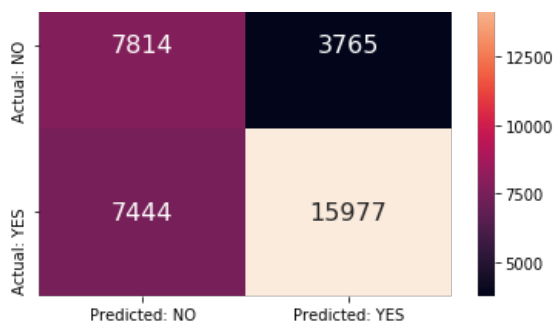
```

Confusion matrix for train data using Avg W2V



Confusion matrix for test data using Avg W2V





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, 192, 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, 683, 696, 700, 713, 722, 724, 731, 736, 743, 749, 754, 767, 774, 792, 796, 801, 806, 807, 809, 819, 821, 828, 838, 843, 868, 870, 875, 877, 880, 886, 887, 891, 894, 897, 898, 901, 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]

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, 1249, 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, 1563, 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, 1885, 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, 2153, 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, 2443, 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, 2666, 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, 2953, 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, 3228, 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, 3526, 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, 3758, 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, 3973, 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, 4221, 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, 4522, 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, 4795, 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, 5118, 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, 5322, 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, 5624, 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, 5889, 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, 6095, 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, 6377, 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, 6605, 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, 6895, 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, 7144, 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, 7449, 7459, 7467, 7473, 7511, 7521, 7523, 7527, 7531, 7544, 7575, 7578, 7580, 7587, 7588, 7592, 7599, 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, 8556, 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)
(1806, 12176)
```

In [0]:

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

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)
```

```
(12176,) <class 'pandas.core.series.Series'>
(12176, 1) <class 'numpy.ndarray'> [[ 7.          ]
 [ 7.          ]
 [101.         ]
 ...
 [ 5.0379881]
 [ 1.1452295]
 [ 0.6411245]]
```

In [0]:

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

fp_features = []
```


12176
124

```
#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
```

[illegible]

In [0]:

1806

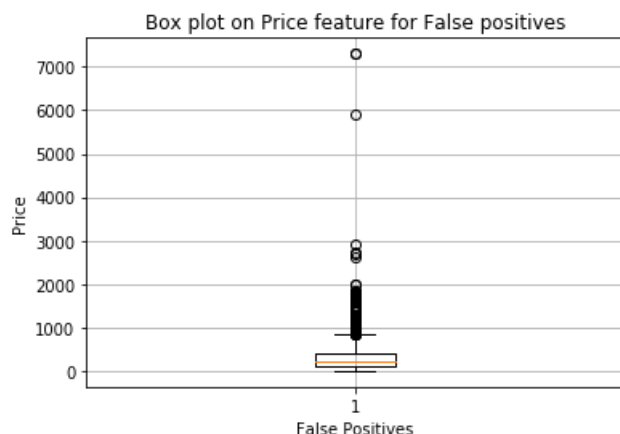
In [0]:

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

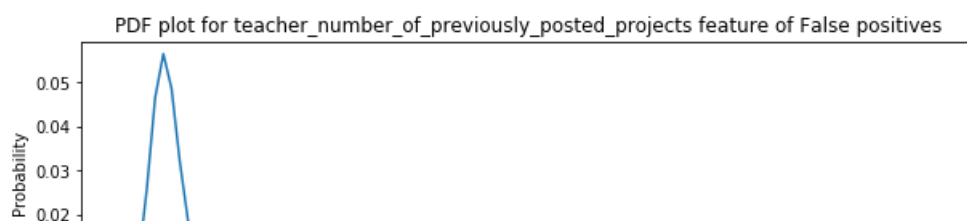
In [0]:

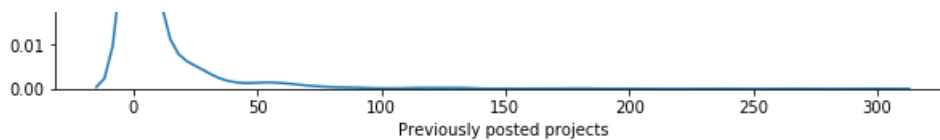
```
#https://www.datacamp.com/community/tutorials/probability-distributions-python

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, 100],
              '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, 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_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)}

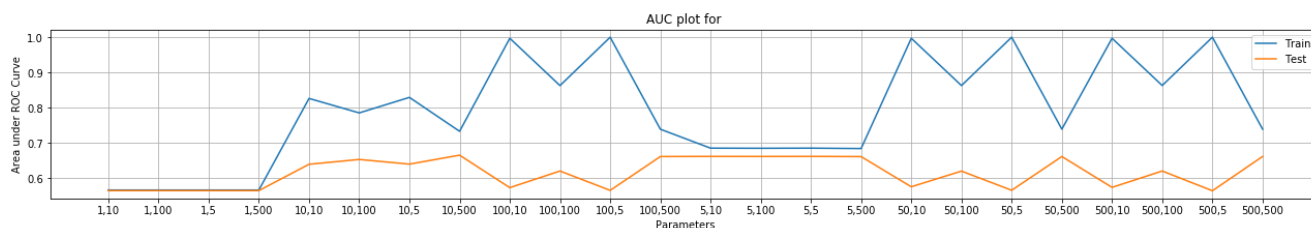
print(params_train)
print(params_test)

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': 0.7385938128715847, '500,5': 0.9998332717860254, '500,500': 0.7385938128715847}

```

```
0.8627068395950728, '500,5': 0.9998362449057137, '500,500': 0.7386429385238374}
{'1,10': 0.5641422578751661, '1,100': 0.5641422578751661, '1,5': 0.5641422578751661, '1,500': 0.56
41422578751661, '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]:

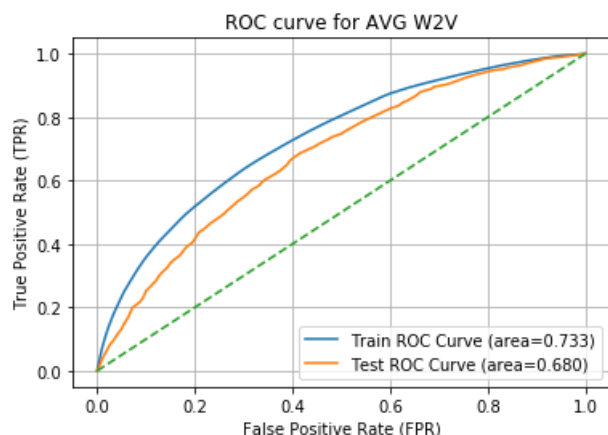
```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#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

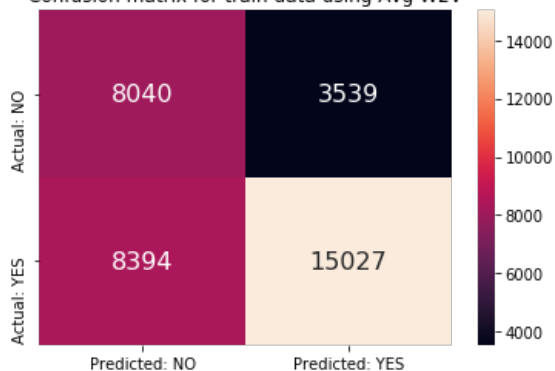
expected_avg_train_w2v = y_train_avg_w2v.values
predicted_avg_train_w2v = final_DT_avg_w2v.predict(x_train_avg_w2v)

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

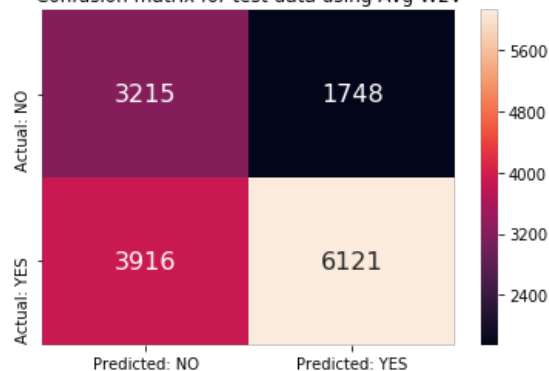
plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_avg_train_w2v, predicted_avg_train_w2v)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train,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

#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)}

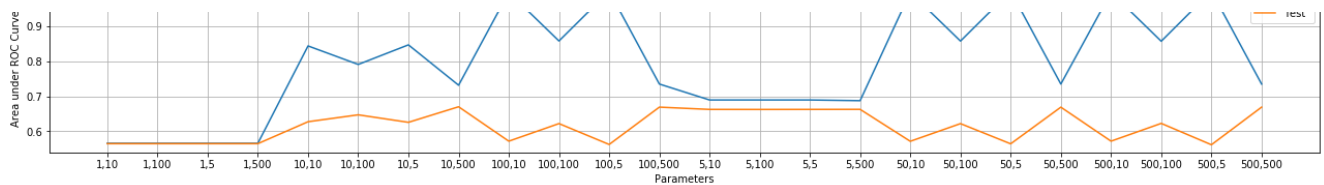
print(params_train)
print(params_test)

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': 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': 0.5648583274471642, '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': 0.562570509233437, '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])
```

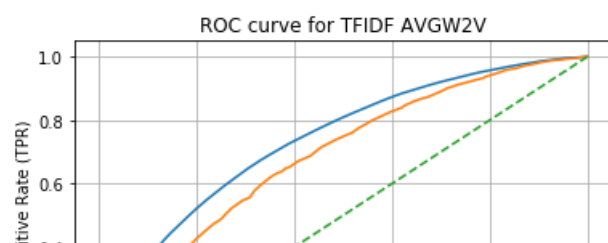
In [0]:

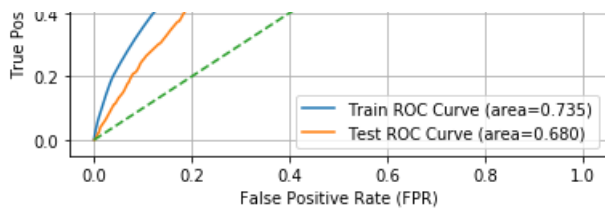
```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#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,
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

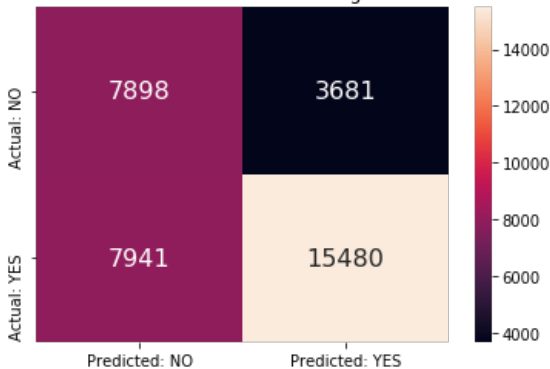
expected_tfidf_train_w2v = y_train_tfidf_w2v.values
predicted_tfidf_train_w2v = final_DT_tfidf_w2v.predict(x_train_tfidf_w2v)

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

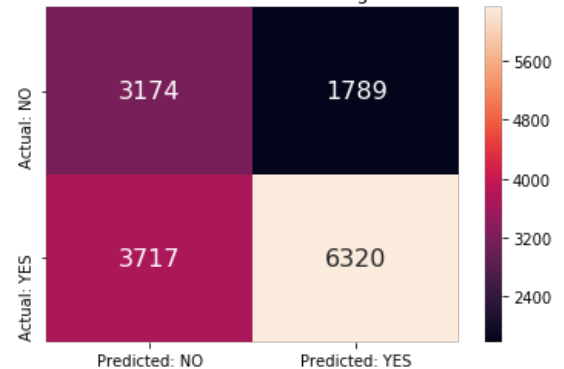
plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_tfidf_train_w2v, predicted_tfidf_train_w2v)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
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

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, 2
4, 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, 9
7, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 11

```

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, 84

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, 2208, 2209, 2210, 2211, 22

0, 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, 3004, 3005, 3006, 3007, 3008, 3009, 3010, 3011, 3012, 3013, 3014, 3015, 3016, 3017, 3018, 3019, 3020, 3021, 3022, 3023, 3024, 3025, 3026, 3027, 3028, 3029, 3030, 3031, 3032, 3033, 3034, 3035, 3036, 3037, 3038, 3039, 3040, 3041, 3042, 3043, 3044, 3045, 3046, 3047, 3048, 3049, 3050, 3051, 3052, 3053, 3054, 3055, 3056, 3057, 3058, 3059, 3060, 3061, 3062, 3063, 3064, 3065, 3066, 3067, 3068, 3069, 3070, 3071, 3072, 3073, 3074, 3075, 3076, 3077, 3078, 3079, 3080, 3081, 3082, 3083, 3084, 3085, 3086, 3087, 3088, 3089, 3090, 3091, 3092, 3093, 3094, 3095, 3096, 3097, 3098, 3099, 3100, 3101, 3102, 3103, 3104, 3105, 3106, 3107, 3108, 3109, 3110, 3111, 3112, 3113, 3114, 3115, 3116, 3117, 3118, 3119, 3120, 3121, 3122, 3123, 3124, 3125, 3126, 3127, 3128, 3129, 3130, 3131, 3132, 3133, 3134, 3135, 3136, 3137, 3138, 3139, 3140, 3141, 3142, 3143, 3144, 3145, 3146, 3147, 3148, 3149, 3150, 3151, 3152, 3153, 3154, 3155, 3156, 3157, 3158, 3159, 3160, 3161, 3162, 3163, 3164, 3165, 3166, 3167, 3168, 3169, 3170, 3171, 3172, 3173, 3174, 3175, 3176, 3177, 3178, 3179, 3180, 3181, 3182, 3183, 3184, 3185, 3186, 3187, 3188, 3189, 3190, 3191, 3192, 3193, 3194, 3195, 3196, 3197, 3198, 3199, 3200, 3201, 3202, 3203, 3204, 3205, 3206, 3207, 3208, 3209, 3210, 3211, 3212, 3213, 3214, 3215, 3216, 3217, 3218, 3219, 3220, 3221, 3222, 3223, 3224, 3225, 3226, 3227, 3228, 3229, 3230, 3231, 3232, 3233, 3234, 3235, 3236, 3237, 3238, 3239, 3240, 3241, 3242, 3243, 3244, 3245, 3246, 3247, 3248, 3249, 3250, 3251, 3252, 3253, 3254, 3255, 3256, 3257, 3258, 3259, 3260, 3261, 3262, 3263, 3264, 3265, 3266, 3267, 3268, 3269, 3270, 3271, 3272, 3273, 3274, 3275, 3276, 3277, 3278, 3279, 3280, 3281, 3282, 3283, 3284, 3285, 3286, 3287, 3288, 3289, 3290, 3291, 3292, 3293, 3294, 3295, 3296, 3297, 3298, 3299, 3300, 3301, 3302, 3303, 3304, 3305, 3306, 3307, 3308, 3309, 3310, 3311, 3312, 3313, 3314, 3315, 3316, 3317, 3318, 3319, 3320, 3321, 3322, 3323, 3324, 3325, 3326, 3327, 3328, 3329, 3330, 3331, 3332, 3333, 3334, 3335, 3336, 3337, 3338, 3339, 3340, 3341, 3342, 3343, 3344, 3345, 3346, 3347, 3348, 3349, 3350, 3351, 3352, 3353, 3354, 3355, 3356, 3357, 3358, 3359, 3360, 3361, 3362, 3363, 3364, 3365, 3366, 3367, 3368, 3369, 3370, 3371, 3372, 3373, 3374, 3375, 3376, 3377, 3378, 3379, 3380, 3381, 3382, 3383, 3384, 3385, 3386, 3387, 3388, 3389, 3390, 3391, 3392, 3393, 3394, 3395, 3396, 3397, 3398, 3399, 3400, 3401, 3402, 3403, 3404, 3405, 3406, 3407, 3408, 3409, 3410, 3411, 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,

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, 4275, 4276, 4277, 4278, 4279, 4280, 4281, 4282, 4283, 4284, 4285, 4286, 4287, 4288, 4289, 4290, 4291, 4292, 4293, 4294, 4295, 4296, 4297, 4298, 4299, 4300, 4301, 4302, 4303, 4304, 4305, 4306, 4307, 4308, 4309, 4310, 4311, 4312, 4313, 4314, 4315, 4316, 4317, 4318, 4319, 4320, 4321, 4322, 4323, 4324, 4325, 4326, 4327, 4328, 4329, 4330, 4331, 4332, 4333, 4334, 4335, 4336, 4337, 4338, 4339, 4340, 4341, 4342, 4343, 4344, 4345, 4346, 4347, 4348, 4349, 4350, 4351, 4352, 4353, 4354, 4355, 4356, 4357, 4358, 4359, 4360, 4361, 4362, 4363, 4364, 4365, 4366, 4367, 4368, 4369, 4370, 4371, 4372, 4373, 4374, 4375, 4376, 4377, 4378, 4379, 4380, 4381, 4382, 4383, 4384, 4385, 4386, 4387, 4388, 4389, 4390, 4391, 4392, 4393, 4394, 4395, 4396, 4397, 4398, 4399, 4400, 4401, 4402, 4403, 4404, 4405, 4406, 4407, 4408, 4409, 4410, 4411, 4412, 4413, 4414, 4415, 4416, 4417, 4418, 4419, 4420, 4421, 4422, 4423, 4424, 4425, 4426, 4427, 4428, 4429, 4430, 4431, 4432, 4433, 4434, 4435, 4436, 4437, 4438, 4439, 4440, 4441, 4442, 4443, 4444, 4445, 4446, 4447, 4448, 4449, 4450, 4451, 4452, 4453, 4454, 4455, 4456, 4457, 4458, 4459, 4460, 4461, 4462, 4463, 4464, 4465, 4466, 4467, 4468, 4469, 4470, 4471, 4472, 4473, 4474, 4475, 4476, 4477, 4478, 4479, 4480, 4481, 4482, 4483, 4484, 4485, 4486, 4487, 4488, 4489, 4490, 4491, 4492, 4493, 4494, 4495, 4496, 4497, 4498, 4499, 4500, 4501, 4502, 4503, 4504, 4505, 4506, 4507, 4508, 4509, 4510, 4511, 4512, 4513, 4514, 4515, 4516, 4517, 4518, 4519, 4520, 4521, 4522, 4523, 4524, 4525, 4526, 4527, 4528, 4529, 4530, 4531, 4532, 4533, 4534, 4535, 4536, 4537, 4538, 4539, 4540, 4541, 4542, 4543, 4544, 4545, 4546, 4547, 4548, 4549, 4550, 4551, 4552, 4553, 4554, 4555, 4556, 4557, 4558, 4559, 4560, 4561, 4562, 4563, 4564, 4565, 4566, 4567, 4568, 4569, 4570, 4571, 4572, 4573, 4574, 4575, 4576, 4577, 4578, 4579, 4580, 4581, 4582, 4583, 4584, 4585, 4586, 4587, 4588, 4589, 4590, 4591, 4592, 4593, 4594, 4595, 4596, 4597, 4598, 4599, 4600, 4601, 4602, 4603, 4604, 4605, 4606, 4607, 4608, 4609, 4610, 4611, 4612, 4613, 4614, 4615, 4616, 4617, 4618, 4619, 4620, 4621, 4622, 4623, 4624, 4625, 4626, 4627, 4628, 4629, 4630, 4631, 4632, 4633, 4634, 4635, 4636, 4637, 4638, 4639, 4640, 4641, 4642, 4643, 4644, 4645, 4646, 4647, 4648, 4649, 4650, 4651, 4652, 4653, 4654, 4655, 4656, 4657, 4658, 4659, 4660, 4661, 4662, 4663, 4664, 4665, 4666, 4667, 4668, 4669, 4670, 4671, 4672, 4673, 4674, 4675, 4676, 4677, 4678, 4679, 4680, 4681, 4682, 4683, 4684, 4685, 4686, 4687, 4688, 4689, 4690, 4691, 4692, 4693, 4694, 4695, 4696, 4697, 4698, 4699, 4700, 4701, 4702, 4703, 4704, 4705, 4706, 4707, 4708, 4709, 4710, 4711, 4712, 4713, 4714, 4715, 4716, 4717, 4718, 4719, 4720, 4721, 4722, 4723, 4724, 4725, 4726, 4727, 4728, 4729, 4730, 4731, 4732, 4733, 4734, 4735, 4736, 4737, 4738, 4739, 4740, 4741, 4742, 4743, 4744, 4745, 4746, 4747, 4748,

1, 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, 5545, 5546, 5547, 5548, 5549, 5550, 5551, 5552, 5553, 5554, 5555, 5556, 5557, 5558, 5559, 5560, 5561, 5562, 5563, 5564, 5565, 5566, 5567, 5568, 5569, 5570, 5571, 5572, 5573, 5574, 5575, 5576, 5577, 5578, 5579, 5580, 5581, 5582, 5583, 5584, 5585, 5586, 5587, 5588, 5589, 5590, 5591, 5592, 5593, 5594, 5595, 5596, 5597, 5598, 5599, 5600, 5601, 5602, 5603, 5604, 5605, 5606, 5607, 5608, 5609, 5610, 5611, 5612, 5613, 5614, 5615, 5616, 5617, 5618, 5619, 5620, 5621, 5622, 5623, 5624, 5625, 5626, 5627, 5628, 5629, 5630, 5631, 5632, 5633, 5634, 5635, 5636, 5637, 5638, 5639, 5640, 5641, 5642, 5643, 5644, 5645, 5646, 5647, 5648, 5649, 5650, 5651, 5652, 5653, 5654, 5655, 5656, 5657, 5658, 5659, 5660, 5661, 5662, 5663, 5664, 5665, 5666, 5667, 5668, 5669, 5670, 5671, 5672, 5673, 5674, 5675, 5676, 5677, 5678, 5679, 5680, 5681, 5682, 5683, 5684, 5685, 5686, 5687, 5688, 5689, 5690, 5691, 5692, 5693, 5694, 5695, 5696, 5697, 5698, 5699, 5700, 5701, 5702, 5703, 5704, 5705, 5706, 5707, 5708, 5709, 5710, 5711, 5712, 5713, 5714, 5715, 5716, 5717, 5718, 5719, 5720, 5721, 5722, 5723, 5724, 5725, 5726, 5727, 5728, 5729, 5730, 5731, 5732, 5733, 5734, 5735, 5736, 5737, 5738, 5739, 5740, 5741, 5742, 5743, 5744, 5745, 5746, 5747, 5748, 5749, 5750, 5751, 5752, 5753, 5754, 5755, 5756, 5757, 5758, 5759, 5760, 5761, 5762, 5763, 5764, 5765, 5766, 5767, 5768, 5769, 5770, 5771, 5772, 5773, 5774, 5775, 5776, 5777, 5778, 5779, 5780, 5781, 5782, 5783, 5784, 5785, 5786, 5787, 5788, 5789, 5790, 5791, 5792, 5793, 5794, 5795, 5796, 5797, 5798, 5799, 5800, 5801, 5802, 5803, 5804, 5805, 5806, 5807, 5808, 5809, 5810, 5811, 5812, 5813, 5814, 5815, 5816, 5817, 5818, 5819, 5820, 5821, 5822, 5823, 5824, 5825, 5826, 5827, 5828, 5829, 5830, 5831, 5832, 5833, 5834, 5835, 5836, 5837, 5838, 5839, 5840, 5841, 5842, 5843, 5844, 5845, 5846, 5847, 5848, 5849, 5850, 5851, 5852, 5853, 5854, 5855, 5856, 5857, 5858, 5859, 5860, 5861, 5862, 5863, 5864, 5865, 5866, 5867, 5868, 5869, 5870, 5871, 5872, 5873, 5874, 5875, 5876, 5877, 5878, 5879, 5880, 5881, 5882, 5883, 5884, 5885, 5886, 5887, 5888, 5889, 5890, 5891, 5892, 5893, 5894, 5895, 5896, 5897, 5898, 5899, 5900, 5901, 5902, 5903, 5904, 5905, 5906, 5907, 5908, 5909, 5910, 5911, 5912, 5913, 5914, 5915, 5916, 5917, 5918, 5919, 5920, 5921, 5922, 5923, 5924, 5925, 5926, 5927, 5928, 5929, 5930, 5931, 5932, 5933, 5934, 5935, 5936, 5937, 5938, 5939, 5940, 5941, 5942, 5943, 5944, 5945, 5946, 5947, 5948, 5949, 5950, 5951, 5952, 5953, 5954, 5955, 5956, 5957, 5958, 5959, 5960, 5961, 5962, 5963, 5964, 5965, 5966, 5967, 5968, 5969, 5970, 5971, 5972, 5973, 5974, 5975, 5976, 5977, 5978, 5979, 5980, 5981, 5982, 5983, 5984, 5985, 5986, 5987, 5988, 5989, 5990, 5991, 5992, 5993, 5994, 5995, 5996, 5997, 5998, 5999, 6000, 6001, 6002, 6003, 6004, 6005, 6006, 6007, 6008, 6009, 6010, 6011, 6012, 6013, 6014, 6015, 6016, 6017, 6018, 6019, 6020, 6021, 6022, 6023,

[illegible]

2, 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, 8082, 8083, 8084, 8085, 8086, 8087, 8088, 8089, 8090, 8091, 8092, 8093, 8094, 8095, 8096, 8097, 8098, 8099, 8100, 8101, 8102, 8103, 8104, 8105, 8106, 8107, 8108, 8109, 8110, 8111, 8112, 8113, 8114, 8115, 8116, 8117, 8118, 8119, 8120, 8121, 8122, 8123, 8124, 8125, 8126, 8127, 8128, 8129, 8130, 8131, 8132, 8133, 8134, 8135, 8136, 8137, 8138, 8139, 8140, 8141, 8142, 8143, 8144, 8145, 8146, 8147, 8148, 8149, 8150, 8151, 8152, 8153, 8154, 8155, 8156, 8157, 8158, 8159, 8160, 8161, 8162, 8163, 8164, 8165, 8166, 8167, 8168, 8169, 8170, 8171, 8172, 8173, 8174, 8175, 8176, 8177, 8178, 8179, 8180, 8181, 8182, 8183, 8184, 8185, 8186, 8187, 8188, 8189, 8190, 8191, 8192, 8193, 8194, 8195, 8196, 8197, 8198, 8199, 8200, 8201, 8202, 8203, 8204, 8205, 8206, 8207, 8208, 8209, 8210, 8211, 8212, 8213, 8214, 8215, 8216, 8217, 8218, 8219, 8220, 8221, 8222, 8223, 8224, 8225, 8226, 8227, 8228, 8229, 8230, 8231, 8232, 8233, 8234, 8235, 8236, 8237, 8238, 8239, 8240, 8241, 8242, 8243, 8244, 8245, 8246, 8247, 8248, 8249, 8250, 8251, 8252, 8253, 8254, 8255, 8256, 8257, 8258, 8259, 8260, 8261, 8262, 8263, 8264, 8265, 8266, 8267, 8268, 8269, 8270, 8271, 8272, 8273, 8274, 8275, 8276, 8277, 8278, 8279, 8280, 8281, 8282, 8283, 8284, 8285, 8286, 8287, 8288, 8289, 8290, 8291, 8292, 8293, 8294, 8295, 8296, 8297, 8298, 8299, 8300, 8301, 8302, 8303, 8304, 8305, 8306, 8307, 8308, 8309, 8310, 8311, 8312, 8313, 8314, 8315, 8316, 8317, 8318, 8319, 8320, 8321, 8322, 8323, 8324, 8325, 8326, 8327, 8328, 8329, 8330, 8331, 8332, 8333, 8334, 8335, 8336, 8337, 8338, 8339, 8340, 8341, 8342, 8343, 8344, 8345, 8346, 8347, 8348, 8349, 8350, 8351, 8352, 8353, 8354, 8355, 8356, 8357, 8358, 8359, 8360, 8361, 8362, 8363, 8364, 8365, 8366, 8367, 8368, 8369, 8370, 8371, 8372, 8373, 8374, 8375, 8376, 8377, 8378, 8379, 8380, 8381, 8382, 8383, 8384, 8385, 8386, 8387, 8388, 8389, 8390, 8391, 8392, 8393, 8394, 8395, 8396, 8397, 8398, 8399, 8400, 8401, 8402, 8403, 8404, 8405, 8406, 8407, 8408, 8409, 8410, 8411, 8412, 8413, 8414, 8415, 8416, 8417, 8418, 8419, 8420, 8421, 8422, 8423, 8424, 8425, 8426, 8427, 8428, 8429, 8430, 8431, 8432, 8433, 8434, 8435, 8436, 8437, 8438, 8439, 8440, 8441, 8442, 8443, 8444, 8445, 8446, 8447, 8448, 8449, 8450, 8451, 8452, 8453, 8454, 8455, 8456, 8457, 8458, 8459, 8460, 8461, 8462, 8463, 8464, 8465, 8466, 8467, 8468, 8469, 8470, 8471, 8472, 8473, 8474, 8475, 8476, 8477, 8478, 8479, 8480, 8481, 8482, 8483, 8484, 8485, 8486, 8487, 8488, 8489, 8490, 8491, 8492, 8493, 8494, 8495, 8496, 8497, 8498, 8499, 8500, 8501, 8502, 8503, 8504, 8505, 8506, 8507, 8508, 8509, 8510, 8511, 8512, 8513, 8514, 8515, 8516, 8517, 8518, 8519, 8520, 8521, 8522, 8523, 8524, 8525, 8526, 8527, 8528, 8529, 8530, 8531, 8532, 8533, 8534, 8535, 8536, 8537, 8538, 8539, 8540, 8541, 8542, 8543, 8544, 8545, 8546, 8547, 8548, 8549, 8550, 8551, 8552, 8553, 8554, 8555, 8556, 8557, 8558, 8559, 8560, 8561, 8562, 8563, 8564,

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, 9357, 9358, 9359, 9360, 9361, 9362, 9363, 9364, 9365, 9366, 9367, 9368, 9369, 9370, 9371, 9372, 9373, 9374, 9375, 9376, 9377, 9378, 9379, 9380, 9381, 9382, 9383, 9384, 9385, 9386, 9387, 9388, 9389, 9390, 9391, 9392, 9393, 9394, 9395, 9396, 9397, 9398, 9399, 9400, 9401, 9402, 9403, 9404, 9405, 9406, 9407, 9408, 9409, 9410, 9411, 9412, 9413, 9414, 9415, 9416, 9417, 9418, 9419, 9420, 9421, 9422, 9423, 9424, 9425, 9426, 9427, 9428, 9429, 9430, 9431, 9432, 9433, 9434, 9435, 9436, 9437, 9438, 9439, 9440, 9441, 9442, 9443, 9444, 9445, 9446, 9447, 9448, 9449, 9450, 9451, 9452, 9453, 9454, 9455, 9456, 9457, 9458, 9459, 9460, 9461, 9462, 9463, 9464, 9465, 9466, 9467, 9468, 9469, 9470, 9471, 9472, 9473, 9474, 9475, 9476, 9477, 9478, 9479, 9480, 9481, 9482, 9483, 9484, 9485, 9486, 9487, 9488, 9489, 9490, 9491, 9492, 9493, 9494, 9495, 9496, 9497, 9498, 9499, 9500, 9501, 9502, 9503, 9504, 9505, 9506, 9507, 9508, 9509, 9510, 9511, 9512, 9513, 9514, 9515, 9516, 9517, 9518, 9519, 9520, 9521, 9522, 9523, 9524, 9525, 9526, 9527, 9528, 9529, 9530, 9531, 9532, 9533, 9534, 9535, 9536, 9537, 9538, 9539, 9540, 9541, 9542, 9543, 9544, 9545, 9546, 9547, 9548, 9549, 9550, 9551, 9552, 9553, 9554, 9555, 9556, 9557, 9558, 9559, 9560, 9561, 9562, 9563, 9564, 9565, 9566, 9567, 9568, 9569, 9570, 9571, 9572, 9573, 9574, 9575, 9576, 9577, 9578, 9579, 9580, 9581, 9582, 9583, 9584, 9585, 9586, 9587, 9588, 9589, 9590, 9591, 9592, 9593, 9594, 9595, 9596, 9597, 9598, 9599, 9600, 9601, 9602, 9603, 9604, 9605, 9606, 9607, 9608, 9609, 9610, 9611, 9612, 9613, 9614, 9615, 9616, 9617, 9618, 9619, 9620, 9621, 9622, 9623, 9624, 9625, 9626, 9627, 9628, 9629, 9630, 9631, 9632, 9633, 9634, 9635, 9636, 9637, 9638, 9639, 9640, 9641, 9642, 9643, 9644, 9645, 9646, 9647, 9648, 9649, 9650, 9651, 9652, 9653, 9654, 9655, 9656, 9657, 9658, 9659, 9660, 9661, 9662, 9663, 9664, 9665, 9666, 9667, 9668, 9669, 9670, 9671, 9672, 9673, 9674, 9675, 9676, 9677, 9678, 9679, 9680, 9681, 9682, 9683, 9684, 9685, 9686, 9687, 9688, 9689, 9690, 9691, 9692, 9693, 9694, 9695, 9696, 9697, 9698, 9699, 9700, 9701, 9702, 9703, 9704, 9705, 9706, 9707, 9708, 9709, 9710, 9711, 9712, 9713, 9714, 9715, 9716, 9717, 9718, 9719, 9720, 9721, 9722, 9723, 9724, 9725, 9726, 9727, 9728, 9729, 9730, 9731, 9732, 9733, 9734, 9735, 9736, 9737, 9738, 9739, 9740, 9741, 9742, 9743, 9744, 9745, 9746, 9747, 9748, 9749, 9750, 9751, 9752, 9753, 9754, 9755, 9756, 9757, 9758, 9759, 9760, 9761, 9762, 9763, 9764, 9765, 9766, 9767, 9768, 9769, 9770, 9771, 9772, 9773, 9774, 9775, 9776, 9777, 9778, 9779, 9780, 9781, 9782, 9783, 9784, 9785, 9786, 9787, 9788, 9789, 9790, 9791, 9792, 9793, 9794, 9795, 9796, 9797, 9798, 9799, 9800, 9801, 9802, 9803, 9804, 9805, 9806, 9807, 9808, 9809, 9810, 9811, 9812, 9813, 9814,

359	10346	10347	10348	10349	10350	10351	10352	10353	10354	10355	10356	10357	10358	10359
373	10360	10361	10362	10363	10364	10365	10366	10367	10368	10369	10370	10371	10372	10373
387	10374	10375	10376	10377	10378	10379	10380	10381	10382	10383	10384	10385	10386	10387
401	10388	10389	10390	10391	10392	10393	10394	10395	10396	10397	10398	10399	10400	10401
415	10402	10403	10404	10405	10406	10407	10408	10409	10410	10411	10412	10413	10414	10415
429	10416	10417	10418	10419	10420	10421	10422	10423	10424	10425	10426	10427	10428	10429
443	10430	10431	10432	10433	10434	10435	10436	10437	10438	10439	10440	10441	10442	10443
457	10444	10445	10446	10447	10448	10449	10450	10451	10452	10453	10454	10455	10456	10457
471	10458	10459	10460	10461	10462	10463	10464	10465	10466	10467	10468	10469	10470	10471
485	10472	10473	10474	10475	10476	10477	10478	10479	10480	10481	10482	10483	10484	10485
499	10486	10487	10488	10489	10490	10491	10492	10493	10494	10495	10496	10497	10498	10499
513	10500	10501	10502	10503	10504	10505	10506	10507	10508	10509	10510	10511	10512	10513
527	10514	10515	10516	10517	10518	10519	10520	10521	10522	10523	10524	10525	10526	10527
541	10528	10529	10530	10531	10532	10533	10534	10535	10536	10537	10538	10539	10540	10541
555	10542	10543	10544	10545	10546	10547	10548	10549	10550	10551	10552	10553	10554	10555
569	10556	10557	10558	10559	10560	10561	10562	10563	10564	10565	10566	10567	10568	10569
583	10570	10571	10572	10573	10574	10575	10576	10577	10578	10579	10580	10581	10582	10583
597	10584	10585	10586	10587	10588	10589	10590	10591	10592	10593	10594	10595	10596	10597
611	10598	10599	10600	10601	10602	10603	10604	10605	10606	10607	10608	10609	10610	10611
625	10612	10613	10614	10615	10616	10617	10618	10619	10620	10621	10622	10623	10624	10625
639	10626	10627	10628	10629	10630	10631	10632	10633	10634	10635	10636	10637	10638	10639
653	10640	10641	10642	10643	10644	10645	10646	10647	10648	10649	10650	10651	10652	10653
667	10654	10655	10656	10657	10658	10659	10660	10661	10662	10663	10664	10665	10666	10667
681	10668	10669	10670	10671	10672	10673	10674	10675	10676	10677	10678	10679	10680	10681
695	10682	10683	10684	10685	10686	10687	10688	10689	10690	10691	10692	10693	10694	10695
709	10696	10697	10698	10699	10700	10701	10702	10703	10704	10705	10706	10707	10708	10709
723	10710	10711	10712	10713	10714	10715	10716	10717	10718	10719	10720	10721	10722	

```
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, 1201
```

[illegible]

.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0007884816580246701, 0.0, 0.0,
0.0005572266655908464, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001593400889645597,
0.0003026386913805423, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0004793368113759322, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.000165392496855456, 0.00011471672139605787, 0.0, 0.0, 0.0006181217528494047, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0025694879193091254, 0.0, 0.001457629197765403, 0.0, 0.0, 0.0004880178463070694, 0.0,
0.0, 0.0, 0.0, 0.0, 7.331199139802395e-05, 0.00022321049383871057, 0.0, 0.0, 0.0, 0.0,
0.0003746971973144906, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00037073901556137486, 0.0,
0.00014412273354363705, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.18
0871445271627e-05, 0.0, 0.0, 0.0, 0.0007788700609838797, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
.0, 0.0, 0.00015825398955787433, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015065374035325458,
0.00023811552433946083, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00022099409557738937,
0.0002874457525068324, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0007053503806364148, 0.0, 0.00034530369304777394, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0008168794547630268, 0.0, 0.0003071595241090076, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00037246941097795913, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00046836024101563857, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.010962213987230797, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00048613416804906885, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000378976428220373, 0.0, 0.0, 0.0001506775600087447, 0.0,
0.00015403659011960314, 0.0002486882060295346, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.00016973802396436598, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.00016851251690690565, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00017172259852018948, 0.0,
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, 0.0009010034620140199, 0.
0, 0.0, 0.0, 0.00016744418432339904, 0.0, 0.0, 0.0, 0.0, 7.889287005683432e-05, 0.0, 0.0, 0.0, 0.0,
, 0.000564201957367773, 0.0, 0.0, 0.0, 6.846853015179721e-05, 0.0004381690337751848, 0.0, 0.0, 8.4
23663474640572e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005642369145761186, 0.0, 8.29229378209341e-05
, 0.0, 0.0, 0.0, 8.428715680178533e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001574015852929195, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0003664609270517343, 0.0, 0.0, 0.0, 0.0, 8.412093196195349e-05, 0.0,
0.0001513575277104387, 0.0, 8.325113903789737e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00019747894560541722, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005095513218532752, 0.0, 0.
.0, 0.00033723749020407107, 0.0, 0.0, 0.000568631983633424, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.331199139802772e-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, 0.0004695005551939664, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0, 0.0, 0.000154123950934975, 0.0, 0.0, 0.00038923124440512247,
0.00020560700021966515, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
8.136699839424753e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003639411528614609, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0001589836357322111, 0.00017591894657419937, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00031815481305821615, 0.0002980832143635104, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00033227873355072627, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005532592246303586, 0.
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.451826907654182e-05,
0.00016357971301081956, 0.0, 8.176618047971895e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0006048038183387407, 0.0, 0.00023970290405949112, 0.00015686212995236348, 0.0,
0.00021572258154059693, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0004000648067026776, 0.0, 0.0, 0.
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0,
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.362835916812515e-05, 0.0
, 0.0, 0.0, 0.0, 0.00044777561369063734, 0.0, 0.0, 0.0, 0.0, 0.00031841139726845327,
0.00015686212995236255, 0.0, 0.0, 0.0, 0.0, 0.0, 8.381919633165516e-05, 0.0, 0.0, 0.0,
0.00046197630412438524, 0.00018363057051246212, 0.0003029674759460917, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.002406127593149056, 0.0, 0.0, 0.0
004662895284836825, 7.331199139802359e-05, 0.001220934220698444, 0.0006546956654020253, 0.0, 0.0,
0.0, 0.0, 0.0008651145344193435, 0.0001668071358336105, 0.00019095448599818542,
0.0001500021641296451, 0.0, 0.0, 0.00031813081195946875, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
8.060671972060288e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000606547218887288, 0.0, 0.0, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.600011400013357e-05, 0.0002353639483306659,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002335779741068639, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0001403930861175955, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00011471672139605066, 0.0, 0.0, 0.0, 0.0, 0.
.0, 0.0, 0.0, 0.00015204210859079675, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0005346921441570409, 0.0, 8.198182136467055e-05, 0.0026772815101742, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.00023047138292554224, 0.00018648640393094834, 0.0, 0.0, 0.0007825227164418389,
0.0, 0.00027419327103416895, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.168144164871691e-
05, 0.00030235830400707415, 0.008734297162693326, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.00014866492472052526, 0.0, 0.0, 0.0, 0.0, 0.0005506119005116136, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0001721518771584846, 0.0009323025429255088, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.004091933332617995, 0.0, 0.0007926470713929366, 0.0, 0.0,
0.0001712979787286365, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00023456732245530286, 0.0036840684200351746, 0.0
005495365037826792, 0.0, 0.0, 0.00044454828481727064, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00015663445715122384, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015541212630908364, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
016778765493879845, 0.0, 0.0, 0.0, 0.0003613169348356491, 0.0, 0.0, 0.0001695658393907826, 0.0, 0.
0, 0.0, 0.0002330398569317647, 0.00016176237242698322, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
.0, 0.0, 0.0, 0.0, 7.600011400015435e-05, 0.000147811466474448, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
.0, 0.0, 0.0, 0.0003998471322952944, 0.0, 0.0, 0.0, 0.00092648750826241, 0.0, 0.00013763895476826427, 0.
00022046957922829718, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.600011400015004e-05, 0.0, 0.0, 0.0,
0.00029426929826948826, 0.0, 0.0, 0.0, 0.00020310409487818488, 0.0, 0.0, 0.0,
0.00015809709960678864, 0.0, 0.0, 0.0, 0.0005628455453753534, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
.00016723111682282154, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0032743672427741587, 0.0, 0.0,

0.0005364836822190728, 0.0, 0.0014966603602280706, 0.0, 0.0, 0.0, 0.0002883461072827349, 0.0, 0.0,
0.0006826772484941922, 0.0, 8.049897893100097e-05, 8.481902176940364e-05, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0005397957316109512, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000470295120519631, 0.0
, 0.0, 0.00020551392196336042, 0.00016539827086221106, 0.0, 0.0, 0.0, 0.00020248749470270686,
0.0018962738289530546, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000192607590425882, 0.0,
6.846853015179881e-05, 0.0, 0.0, 0.00016959517642977235, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00024737002262410686, 0.0, 0.00037838246892259305, 0.0, 0.00015986734284165066, 0.0, 0.0, 0.0, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003262475762332625, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0002782920878779743, 0.00016153889557453697, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0006096158897119557, 8.293560887325106e-05, 0.0, 0.00016780726970268945, 0.0003304080865997181,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0018832651674981916, 0.0, 0.0012992993796419991, 0.0,
7.96777695374278e-05, 0.0, 0.0, 0.00016751368457939103, 0.0, 0.000665507294376367, 0.0, 0.0, 0.0,
0.0, 0.00027644011437452957, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.269559062337499e-05, 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,
0.0002199774762520543, 0.0002677591029564977, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001508345887646957, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001713121391081776, 0.0, 0.0,
0.00018957783509871562, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.353034927216255e-05, 0.0, 0.0,
0.0, 8.413302195613307e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003469055930343838, 0.0, 0.0,
0.00042784344055366626, 0.0, 0.0, 0.0, 0.00015399579066978036, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00045062756298572845, 0.0, 0.0, 0.0, 0.0, 0.0003145109607458415, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00015778574011366864, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0019337037991149267, 0
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00036657545604731965, 0.0, 0.0,
0.0, 0.0, 0.0002427929665245503, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016942027251211177, 0.0, 0.0, 0.0, 0.0,
.0, 0.0002726952669897929, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016952829477827057, 0.0, 0.
0, 0.0, 8.17869026277583e-05, 0.0, 0.0, 0.0, 0.0, 0.0002972032369248478, 0.0, 0.0, 0.0, 0.0,
0.00016638610393502923, 0.0, 0.0, 0.0002989815734117256, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00034524614188419396, 0.0, 0.0, 0.0, 0.0, 0.00028204029728436754, 0.0, 0.0,
0.0003205482061266015, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00031354709499424496, 0.0, 0.0, 0.0, 0.00
020597386590937873, 0.0, 0.0, 0.0, 0.0004066629162100039, 0.0, 8.483778381513485e-05, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.00015759051076478663, 0.0, 0.0, 0.00046513918719759563, 0.0,
0.0002029997494157341, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
7.331199139802782e-05, 0.0002912026335820452, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.001183350006048508, 0.0, 0.0, 0.0, 0.003496522657966758, 0.0, 0.0,
0.0021879607779489254, 0.0003013591974781027, 0.0, 0.0005293196227197007, 0.0, 0.0,
8.429673782712638e-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, 0.0008063819983152562, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016988655509371883, 0.0, 0
.00016136288635555638, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.08954786523158e-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, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00023743639948562066, 0.00016571667679308061, 0.0004936305456012455, 0.0, 0.0010438076753696004,
0.0, 0.0, 0.0, 8.266876829042903e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0010166985596103155,
8.1148682008477e-05, 0.0, 0.0, 0.0, 0.0009491673048371375, 0.0011757758642080129, 0.0, 0.0,
0.0004778529432681816, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0013605842827890558, 0.0, 0.0, 0.0, 8.508390459540394e-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, 0.0006996770590263516, 0.0, 0.0, 0.0001947560094703549, 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, 0.0001530841848339278, 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, 0.0, 0.0, 0.0, 0.0, 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, 0.0003044474010092792, 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, 0.00014801936837289577, 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, 0.00016304238772130758, 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, 0.0003077226301710649, 0.0, 0.0, 0.0
0012297702789112327, 0.0, 0.0, 0.0, 0.001439227145632343, 0.0007706890035443536, 0.0, 0.0,
0.00038726155094298255, 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, 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, 0.0003422673577965711, 0.0, 0.0019469253774606848, 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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00
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, 0.00015306533061748226, 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,
0.0001527313444748711, 0.0008804538165383852, 0.0, 0.0, 0.0, 0.00025669071053778444, 0.0, 0.0, 0.0

[illegible]

[illegible]

0, 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, 0.0, 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, 0.00015837757585855872, 0.0, 0.0, 0.00029712959981112853, 0.0, 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, 0.0, 0.0, 8.3970053467338e-
05, 0.0, 0.0, 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, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0001638023945243879, 0.0, 0.0, 0.0, 0.0, 0.0002930536816576221, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.00020362492997626604, 0.0, 0.0, 0.0, 0.0, 8.12277567240588e-05, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.00017058337394834693, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00022694472120166336,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015507499274074248, 0.0, 0.0, 0.0, 8.043717432437456e-05, 0.0, 0.
.0, 0.0, 0.0005153300223065924, 0.0, 0.0, 0.0, 0.0006724507812564328, 0.0, 6.846853015179721e-05,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0006316416587089352, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0,
, 0.0, 0.0, 0.00014805560850403266, 0.00018128187378406812, 0.0, 0.00040921781832678277,
8.042342935552222e-05, 0.0, 0.0015653585427739022, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0008525111280121505,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005482424076933014, 0.00016965733713886295, 0.0,
0.00018434463348037416, 0.0, 0.0, 0.0001423144543593228, 0.0, 0.0, 0.0001291781043112488,
0.0008589670690681082, 0.0008038099706625652, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.000335604368528892, 0.0, 0.00028189088817655, 0.0, 0.0,
0.00014589826911592585, 0.00033240192362611197, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0,
, 0.0, 0.0, 0.0005131767154619901, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016952829477827206, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 8.394520358268936e-05, 0.0, 8.068474718686291e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 8.406404938952208e-05, 0.0, 0.0, 0.0009889874222294356, 0.0, 7.703974047809466e-
05, 0.0, 0.0, 0.0, 8.325879223651013e-05, 0.0, 0.0, 0.0, 0.0, 0.0005468155264212593, 0.0, 0.0, 6.8
46853015179721e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005329919836621613, 0.0, 0.0, 0.0, 0.0, 0.0,
8.247001388325113e-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,
0.0002223830003147629, 0.0, 0.00020972029419507812, 0.00046987241159033845, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.000568342710891511, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0047677930668384204, 0.00021279304086450917, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005798230517926547,
0.0, 0.0001515789127066382, 0.0, 0.0, 0.0, 0.0, 0.0003628730690279996, 0.00014366680793896554, 8.5
31026265630316e-05, 0.00016419619420912895, 0.0, 0.0, 0.0, 0.0006316139254554168,
0.00015464999476994096, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005713327899026984,
0.0001142857142856666, 0.0003411140467584983, 0.00036676199203530704, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0004635977170499907, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.382776706411535e-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, 0.0, 0.0, 0.0, 0.0,
0.00015300858261639225, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016876088203160748, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00014977435421245852,
0.00015670129527859831, 0.0, 0.0, 0.0, 0.0, 0.0, 0.002365918537102489, 0.00481801622307083, 0.0, 0.
.0, 0.0,
05674530900737071, 0.0, 0.0, 0.0, 0.0, 0.0003352385613278486, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
.0, 0.0,
0.0, 0.0, 0.0, 0.0006317705409035677, 0.0002467972206938564, 0.0, 0.0005789347808908382, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00014781146647444758,
0.0003003093693047917, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.00030763245752343717, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.001241888878597617, 0.0, 0.0, 0.0, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00019933273648216386, 0.0, 8.417745218654465e-05, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00020962554602701452, 0.0, 0.00034801168627563876, 0.0,
0.0018270038694818127, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0009060544789033665, 0.0, 0.0003405586073904358,
0.0, 0.0, 0.0, 8.202849811184431e-05, 0.0, 0.0, 0.0003741952485785338, 0.0, 0.0,
0.0002912046922358688, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00047999046872753185, 0.000292482950819088,
0.0, 0.0003742369189430848, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0013672680942462003, 0.00016726942112632424, 0.0, 0.0002941487213675677, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0, 0.0, 0.000302409044962492, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0002664761270140745, 0.0, 0.0004875780666623954, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 0.0, 8.293560887325106e-05, 0.0, 0.00026229514296956355, 0.0,
0.00019996295414582553, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0009532900536378566,
0.0005040125724500896, 0.0, 0.0, 0.0, 0.00014964859303593088, 8.137060649578034e-05,
7.331199139802782e-05, 0.0, 7.498070540455262e-05, 0.0, 0.0, 0.0, 0.0006033796597465799, 0.0, 0.0,
0.0, 0.0, 0.0, 8.309552500190355e-05, 0.0, 0.0002260257546139227, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0001229770278911214, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 0.0, 0.0007274926708112526, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.005560071987048242, 0.0003212575752584118, 0.0, 0.0020199312099821428,
0.005318483989211942, 0.0, 0.0, 0.0002084917363848821, 0.0002036927088885706, 0.0,
0.00024095053816174043, 0.0, 0.0, 0.0, 0.0003862414980764193, 0.0, 0.0005452183493777387, 0.0, 0.0,
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00045819008352756743, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0.

0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00023098667525748998, 0.0, 0.0, 0.0, 0.0006786677938117235, 0.0
002020509956004747, 0.00014458091278160426, 0.00044655287268356264, 0.0006457415294269936, 0.0,
0.0, 0.0015412706600615265, 7.972062610470122e-05, 0.0, 0.0009980969427593775,
0.00017101027715222397, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005634292302554295, 0
.0, 0.0, 0.0, 0.0, 0.0, 0.0018906358348347392, 0.0, 0.00034930281861691244,
0.00017018694946472905, 0.0003105917158278928, 0.00032107398933864953, 0.0, 0.0,
0.00034838012667374633, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0012367532193996783, 0.0, 0.0,
0.0005041010218095597, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0010345915350956098, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00065517125829789, 0.0, 0.0, 0.0, 0.000881042962216302, 0.0, 0.0, 0.0, 0.0003609682223272467, 0.
0, 0.0, 0.0, 0.0, 0.00041523529323243607, 0.0004236922984183872, 0.0, 0.0, 0.0,
0.0014645995530205174, 0.0, 0.0, 0.0, 0.0011325391738432667, 0.0, 0.000898685589531115, 0.0, 0.0,
0.0, 0.0, 0.0006271087204379571, 0.0002907446909840861, 0.0, 0.0006609710958604016, 0.0,
0.0002614323754110033, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000288998660705256, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015754856581296333, 0.0, 0.0, 0.0,
0.00021153815046079833, 0.0, 0.0009560351611073034, 0.0001538452432613304, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 7.976026343672112e-05, 0.0003375800795338954, 0.0,
0.00022839687974445915, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002472505941468498, 0.0, 0.0, 0.0,
0.0026657836981005658, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00022074374628787904, 0.0,
0.0016997507887624552, 0.0009070722197863982, 8.288152280398921e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0016183052281622094, 0.0003161282014875541, 0.0, 0.0, 0.0, 0.0,
0.00030610973240187135, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003638380371250277, 0.0, 0.0,
0.0004055289064215811, 0.0, 0.0, 0.00022858624696807581, 0.0, 0.0, 0.000723818663841926, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0008395447774224705, 0.0, 0.000158402703474631, 8.039145401045253e-05, 0.0,
0.0, 0.0, 0.0, 0.000234310454493022, 0.0, 0.0, 0.0, 0.0, 8.226494374339883e-05, 0.0, 0.0, 0.0, 0.0
, 0.0, 0.0010887390383388627, 0.0, 0.0029093970959546698, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0003746893309834429, 0.0, 0.0, 0.0, 0.0007460984011408293, 0.0, 0.00026990654836969377, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.255199971045349e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
, 0.0, 0.0, 0.00023408605003189236, 0.0, 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, 0.0031338924067340233, 0.0, 0.0, 0.0002547091584808867, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.001937406220567796, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
.0, 0.0, 0.00032243139495197155, 0.0, 0.0, 0.0, 0.00043499738167922166, 0.006853975533842708, 0.0,
0.001831933630040048, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000357512287650908, 0.0003063971141196238, 0.
0, 0.0, 0.0, 0.00013786804304923692, 0.0, 0.0, 0.0, 0.00015693855813160587, 0.0, 0.0, 0.0, 0.0,
0.0007566763729978896, 0.0, 0.0, 0.0, 8.145499152180823e-05, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00014709667929240004, 0.00036079563063639604, 0.0, 0.00019618659626015252, 0.0, 0.0,
0.00029680203419642557, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.314569256667718e-05, 0.0, 0.0, 0.0, 0.0, 0.
.0, 0.0, 0.0, 0.0, 0.00021760131111292936, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00029586704399258996, 0.00020515792048775785, 0.00046757116687833094, 0.0004012680674406833,
0.0, 0.0003157475414830319, 8.266876829042903e-05, 0.0, 0.0, 0.0, 8.390613850967219e-05,
0.0002490834187770539, 0.0009456653105767626, 0.0, 0.0, 0.0, 8.47790952188534e-05, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005272450552223011, 0.0, 0.0, 0.0, 0.00028735319540610076,
0.0003823332050852934, 0.0, 0.0, 0.0, 0.0, 8.318884027570209e-05, 0.00016559068828763153, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00029899630130676214, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00013786804304919556, 0.00025754143738277016, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00012917810431125365, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.198182136466807e-05, 0.
.000273238464819362, 0.0, 0.0, 8.111213370417701e-05, 0.0, 0.0, 0.0019759760166834016, 0.0, 0.0,
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,
0.000155432356824622, 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,
0.00020222317787422725, 0.00036977398528385374, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
7.77097385840566e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002905853205612918, 0.0, 0.0,
8.278598670995642e-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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 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, 0.00021892959067974975, 0.0,
0.0002529619046366924, 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,
0.00014531690192532107, 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, 0.001436400708625175, 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,
0.00024316749798798228, 0.0001627412129915617, 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, 8.16899497405452e-05,
0.0, 0.0, 0.0, 8.227395246168361e-05, 0.0, 0.0, 0.0005665672500505483, 0.0, 0.0, 0.0, 0.0,
0.0002801651773964964, 0.0, 0.0, 0.0007379020339579158, 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, 0.01243410915485557, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 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,
0.00023383864468982238, 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,
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,
0.0002461966080769925, 0.0, 0.0, 0.0, 6.846853015179721e-05, 0.0, 0.0, 0.0,

[illegible]

[illegible]

0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005921805293213215, 0.0, 0.0,
0.0, 0.0, 0.0, 0.00024001592426002116, 0.0, 0.0, 0.0, 0.0, 0.0006539886005580019,
0.00014589826911592585, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00029539034682449775, 0.0,
0.00036445633565696, 0.0, 0.0, 0.0, 0.004489577828965279, 0.0, 0.0, 0.00016606347836058455,
0.0007710576134345346, 0.0, 0.0, 0.002014384587706046, 0.0003115516853176177,
0.000777670067851175, 0.0, 0.0, 7.331199139802604e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
0023973580466021095, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00027434536000045236, 0.0, 0.0, 0.0, 0.00016849297610468461, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005322106980924333, 0.0,
0.0010754676417096894, 0.0, 0.0, 0.0003826715271660365, 0.00024992192690475456, 0.0, 0.0, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0018405330519257015, 0.0, 0.0, 0.0, 0.0,
0.00043157503108334723, 0.0, 0.00030300781992216683, 0.0, 0.00019839125275018273, 0.0, 0.0, 0.0, 0.0,
.0002770849946545657, 0.0, 0.0, 0.0, 0.0, 8.09468988343296e-05, 0.0, 0.0, 0.0, 0.0,
0.00029047158652514745, 0.0, 0.0, 0.0, 0.0, 0.0014089281573355416, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0
.0, 0.0, 0.0, 0.0, 0.0, 7.770973858440575e-05, 0.0, 0.0, 0.0, 0.0, 0.000212084584248835, 0.0, 0.0, 0.0,
0.0, 0.0, 0.00040615259645337287, 0.0, 0.0, 0.0, 0.00015319722442781015, 0.0, 0.0, 0.0, 0.0, 0.0,
.0, 0.0, 0.000594030873426411, 0.0, 0.0, 0.0, 0.00022383336671980033, 0.0, 0.0002044612815526574,
0.00036672556146037606, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00013266655514417678, 0.0002494602672970889, 0.
0018191965993783464, 0.0, 0.0002436321408504165, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0003607360784577068, 0.0, 0.0, 0.0, 0.0007332239367872386, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0003028011083638105, 0.0,
0.00014589826911592574, 0.0, 0.0013503844077180944, 0.0, 8.29802604603581e-05,
0.00021010671663386, 0.0, 0.0, 0.0, 0.0, 0.0002277499207383047, 0.0, 0.0, 0.0003205335296893515, 0.0, 0
.0, 0.0, 8.429673782712638e-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,
0.00014223957516403273, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00016581389937732726, 0.0, 0.0,
0.0, 0.0, 0.0003839357580658186, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.419328230608594e-05,
8.288689107767236e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0006125675857040539,
0.0003743048351035935, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00022943344279210934, 0.0, 0.0, 0.0, 0.0, 0
.0, 0.0, 8.149665709404109e-05, 0.0, 0.00016337099944285115, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 8.130006280937856e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 8.207915613981234e-05, 0.0, 0.0,
0.0, 0.0
, 0.0001569469445674097, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005317229863235122,
0.0, 0.00024948279342228315, 0.0, 0.0, 0.000358114635707107, 0.0, 0.0, 7.770973858439796e-05, 0.00
04221096921476052, 8.449090967379976e-05, 0.0, 0.0, 0.0, 0.0, 0.0002840786071513388, 0.0, 0.0, 0.0
, 0.0, 0.0, 0.0, 0.0, 0.0004574807996464365, 0.0, 0.00013400463273153455, 0.0014727505845910577, 8
.521643452643941e-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, 0.0,
0.0008055214280061485, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00047336229802304056, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 8.224445311599764e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0012755013040001923, 0.0, 0.0, 0.
0, 0.000301842858451738

[illegible]

0.2400700000000000e-05, 0.0, 0.0, 0.0, 0.0070000000000000e-05, 0.0020000000000000e-05, 0.0, 0.0, 0.0,
0.0, 0.002938657437994217, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00040826057384897823, 0.0, 0.0, 0.
00013721785047189435, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00028584203550819753, 0.0, 0.0,
0, 8.062718816738301e-05, 0.0, 0.001199268178913624, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0011676688008159105, 0.0005838328570499367, 0.0, 0.0, 0.00014750401901987225, 0.0, 0.0, 0.0, 0.0,
, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
6.846853015179721e-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,
0.0006282893322222137, 7.77097385844008e-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, 8.415800787048596e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.000302897434829609, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0002055007116769975, 0.0, 0.0, 0.0, 0.0002563053100953318, 0.0, 0.0, 0.00014903790242007766, 0.0,
, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00030922336991976893, 0.0, 0.0,
0.00028296431391933396, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0051566392369810175, 0.0005785450115746726, 0.00029309510299986563, 0.0, 0.00025258083432381395,
0.0, 0.0, 0.0, 0.0023811238103674498, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0001537729719339969, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000307621245403021, 0.0, 0.0, 0.0, 0.0,
0.0005149579241134468, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.073032533709901e-05, 0.0, 0.0, 0.0, 0.
0021347580536764466, 0.000702719490686606, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.000
1561075005545248, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015076898680877862, 0.0, 0.0, 0.0, 0.
.0, 0.0, 0.0, 0.0008702914190628754, 0.0, 0.00015930067005837426, 0.0004600418870482202,
0.0002097824207490093, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00041074130946866885, 0.0, 0.0,
0.00024151911817526953, 0.0, 0.0014142079356361356, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
8.035123766688025e-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, 0.0001699759857926572, 0.0, 0.0004057917728728028, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005498675423048797, 0.0, 0.
0001436668079389655, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00022078409636363836, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.322783839778426e-05, 0.0, 0.0, 0.
.0, 0.00026642579065657915, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005114406368421759, 0.0016824683319716313, 0.0,
0.00015313279180248547, 0.0, 0.0, 0.0, 0.0, 0.006590639431097807, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.
00021485864447147842, 0.0, 0.0006509303980274571, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00020973263871380818, 0.0, 0.0, 0.0, 0.00041421458144390365, 0.0002462489607424585,
0.0042374264991149975, 0.0, 0.0, 0.0, 0.0, 0.0, 8.42638520668528e-05, 0.0016643666183460237, 0.0,
0.0, 0.0, 0.0, 0.00029953664242384685, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0021348362788930933, 0.
.0, 0.0, 0.0, 0.0, 0.0, 8.344433500553951e-05, 0.0007993637987591279, 0.0, 0.0001465572838134674,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0030765982752689863, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0006840331793671876, 0.0, 0.0, 0.0, 0.00222549327397638, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.002662875415814181, 0.0, 0.0, 0.0, 0.0, 0.0, 8
.172058987404041e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0005190427048835168, 0.0, 0.0,
, 0.00017916734808655116, 0.0, 0.0, 6.846853015179881e-05, 0.000134756584349316, 0.0, 0.0, 0.0, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002474771794124135, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0,
0.0, 7.77097385844223e-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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 9.110148911971579e-
05, 0.0, 0.0001563556585845394, 8.504596699369486e-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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00016188503541739082, 0.0, 0.0, 0.0, 0.0, 0.00020139559029288452, 0.0, 0.0001503188385787059, 0.
0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0002863326464516474, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.00016832559788062526, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.09045626595928e-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, 0.0, 0.0, 0.0, 0.0,
0.0,
, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.960093686120248e-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, 0.0001708704824683828, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
, 0.0, 0.0, 0.0003707950396100934, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0002942556436762, 0.0,
, 0.0008200790008535622, 0.0, 0.00016000114537428252, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 8.426654819312473e-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,
0.00014669234663218132, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0,
, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0,
0.0,
, 0.0,
.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8.32
6130986034833e-05, 7.997029085200305e-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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.331199139802791e-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, 0.00023043256345267268, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00012297702789112514, 0.0,
0.0, 0.0, 0.0, 0.00014034976508747027, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00029827886280438465,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 7.889287005682523e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 8.16961285838089e-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, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.00015938106455697166, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0004933537457142056, 0.0, 0.0, 0.0, 8.386227963949687e-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, 0.00017187439772866088, 0.0, 0.0, 0.00024106694584458425, 0.0,
8.213808438264373e-05, 0.0, 8.351002759881591e-05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.

In [0]:

```
#https://www.geeksforgeeks.org/how-to-drop-one-or-multiple-columns-in-pandas-dataframe/
```

```
print(type(res.keys()), res.keys(), '\n', x_train_set5.shape)
```

```
loc=list(res.keys())
```

```
print(loc)
```

```
df_train_set5=df_train_set5.iloc[:, loc]
```

```
df_test_set5=df_test_set5.iloc[:, loc]
```

```
print(df_train_set5.shape, df_test_set5.shape)
```

```
<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, 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, 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, 2

339, 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, 2
571, 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, 2
829, 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, 3
043, 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, 3
249, 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, 3
462, 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, 3
645, 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, 3
782, 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, 3
958, 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, 4
085, 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, 4
279, 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, 4
485, 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, 4
681, 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, 4
934, 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, 5
114, 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, 5
399, 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, 5
559, 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, 5
698, 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, 5
835, 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, 6
002, 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, 6
208, 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, 6
354, 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, 6
564, 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, 6
760, 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, 6
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, 7
185, 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, 7
369, 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, 7
575, 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, 7
824, 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, 8
106, 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, 8
374, 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, 8
542, 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, 8
775, 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, 8
991, 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, 9
144, 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, 9
327, 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, 9
487, 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, 9
660, 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, 9
968, 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),
             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'])

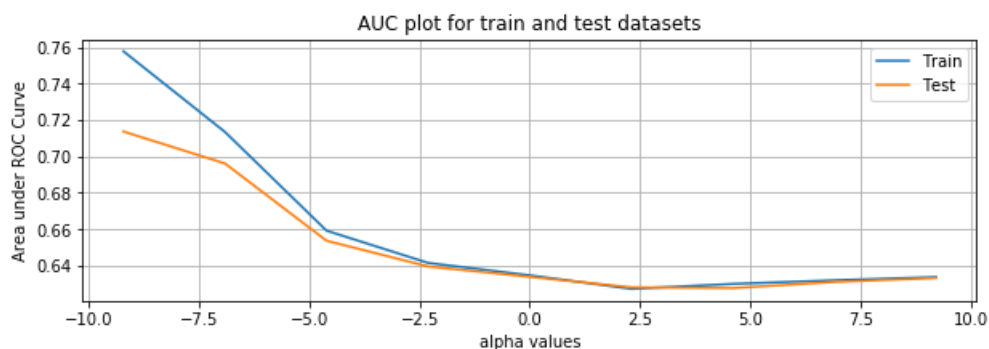
log_params = []
for i in parameters['alpha']:
    log_params.append(math.log(i))

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.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-dire
ctly

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='balan
ced')
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] give
s 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]:

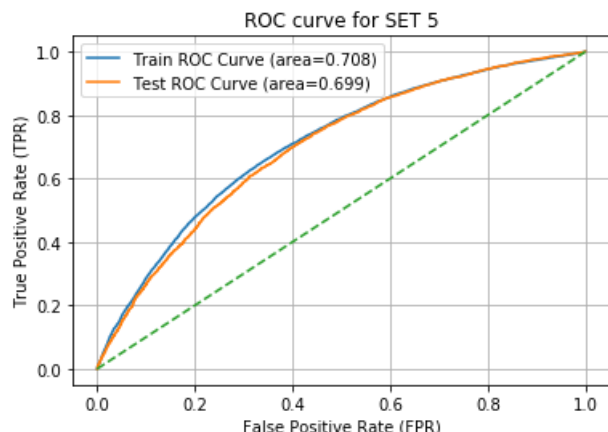
```
#https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
#https://www.programcreek.com/python/example/81207/sklearn.metrics.roc_curve
#https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

#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

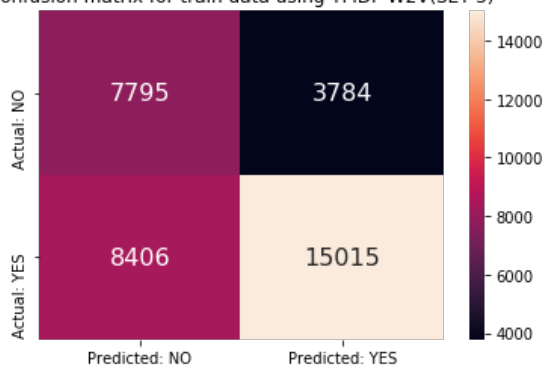
expected_set5_train = y_train_set5.values
predicted_set5_train = final_SGD_set5.predict(final_df_train_set5)

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

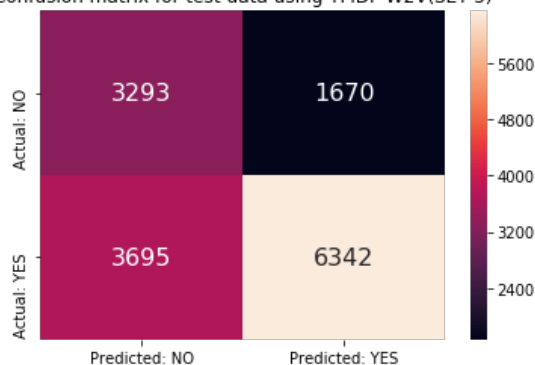
plt.subplots(figsize=(15,4))
plt.subplot(1,2,1)
cmdf_train=cf_mx(expected_set5_train, predicted_set5_train)
df_cm_train = pd.DataFrame(cmdf_train, range(2),range(2))
df_cm_train.columns = ['Predicted: NO','Predicted: YES']
df_cm_train = df_cm_train.rename({0: 'Actual: NO', 1: 'Actual: YES'})
sns.heatmap(df_cm_train,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

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

In [0]: