

ahmed_999177037_assignment1

November 1, 2019

```
[1]: #Importing all the required libraries
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from matplotlib import pyplot as plt
from random import sample
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
from sklearn.model_selection import cross_val_score, cross_val_predict
from sklearn import metrics
from sklearn.model_selection import KFold
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import learning_curve
from sklearn.model_selection import ShuffleSplit

pd.set_option('display.max_rows',500)
pd.set_option('display.max_columns',500)
pd.set_option('display.width',1000)
```

1 DATA CLEANING

In the data cleaning process, I initially encountered a lot of missing values. However, upon further analysis I found out that it was because the survey respondent didn't select the given option from multiple choice question. The following are the steps I took to clean the data:

1. Removing naan records (entire rows) from columns that have less than 10% naan values because the feature column already has a rich amount of data.
2. Removing 'Other Text' column as it has no meaning in the dataset.
3. Identifying features that have more than 70% naan values and removing them.

4. Removing few identified features from the gender and qualification question as it had very low percentage of records (based on data exploration).
5. Cleaning target feature by removing naan records and answers that very low percentage of records selected.
6. All the categorical data with binary text answers are label encoded (0 or 1).
7. All the categorical data with multiple answers are one hot encoded.
8. All the categorical data in the survey that required text answers (which were eventually converted to integers in csv) are converted to either 1 or 0. If the surveyor didn't type anything then the csv file shows -1 which is converted to 0. If the surveyor typed anything then the csv file shows a positive number based on the number of characters, this is converted to 1 which shows the user typed something.
9. All the remaining features with integer data are normalized and the time to complete feature is dropped as it does not add any value.

After the cleaning and encoding process, I ended up with a clean dataset that had 11,390 records and 64 features.

```
[2]: #####PREPROCESSING PART I#####
      ↳1#####

#Importing the file
response=pd.read_csv("multiple_choice_response.csv", encoding = "ISO-8859-1")

#Deleting column header and replacing it with row 1. Also changing column index
response.columns=response.iloc[0]
response=response.drop(response.index[0])
response=response.reset_index(drop=True)

                                ##      DATA CLEANING PART I      ##

###Removing Naan records from column that has less than 10% naan values###
###Why: Because features already has a rich source of data###

#Percentage of responses required for keeping the rows
percentage_required=10

#Determining cutoff number:
total_record=response.shape[0] #total number of records
cutoff_num=round(total_record*(percentage_required/100))

#Identifying features that has less than 10% Naan values
null_records_remove_1=response.isnull().sum(axis=0)
null_records_remove_1=pd.Series(null_records_remove_1).where(lambda x :↳
↳x<cutoff_num).dropna()
```

```

null_records_remove_1=pd.Series(null_records_remove_1).where(lambda x : x!=0).
↳dropna()
rows_remove=null_records_remove_1.index.tolist()

#Removing identifying rows:
response_1=response.dropna(axis=0, subset=rows_remove)

                                ## DATA CLEANING PART II ##

#Removing 'Other Text' column as it has no meaning in the dataset
response_1=response_1.drop(response_1.filter(regex='- Other - Text').columns,
↳axis=1)

#Removing columns that is not important (from observation)#
#Checking which columns have high percentage of Naans -

#Identifying cutoff percentage:
cutoff_percent=70

#Identifying the number of records that resembles that percentage
remaining_records=response_1.shape[0]
cutoff_val=round(remaining_records*(cutoff_percent/100))

#Identifying features that has more than 70% Naan values
null_records_remove_2=response_1.isnull().sum(axis=0)
null_records_remove_2=pd.Series(null_records_remove_2).where(lambda x :
↳x>cutoff_val).dropna()
features_remove_2=null_records_remove_2.index.tolist()

#Removing identified columns:
remov_column=['What is your gender? - Prefer to self-describe - Text']
response_2=response_1.drop(remov_column, axis=1)

#Removing gender records that have very less percentage: [For more information,
↳refer to exploratory data anlaysis]
response_2=response_2[response_2['What is your gender? - Selected Choice'] !=
↳'Prefer to self-describe']
response_2=response_2[response_2['What is your gender? - Selected Choice'] !=
↳'Prefer not to say']

#Removing records where degree is "I prefer not to answer" [For more
↳information refer to exploratory data anlaysis]
response_2=response_2[response_2['What is the highest level of formal education,
↳that you have attained or plan to attain within the next 2 years?'] != 'I
↳prefer not to answer']

```

```

####Removing records that has less than 20 non-Naan values
response_2= response_2.dropna(axis=0,thresh=20)

# CLEANING TARGET FEATURE
#Removing salary (target) that does not provide any information
response_3 = response_2[response_2['What is your current yearly compensation,
↳(approximate $USD)?'] != 'I do not wish to disclose my approximate yearly
↳compensation']

#Removing Naan records from target
response_3 = response_3.dropna(axis=0,subset=['What is your current yearly
↳compensation (approximate $USD)?'])

#Need backup to do further analysis
response_4=response_3.copy()
response_5=response_3.copy()

#####PREPROCESSING PART
↳II#####

#LABEL ENCODING

#Replacing all the Naan values with 'NaN'
response_4[pd.isnull(response_4)] = 'NaN'

label_encode_col=response_4

drop_cols_oh=['What specific programming language do you use most often? -
↳Selected Choice',
               'What programming language would you recommend an aspiring data
↳scientist to learn first? - Selected Choice',
               'Of the choices that you selected in the previous question,
↳which ML library have you used the most? - Selected Choice',
               'Of the choices that you selected in the previous question,
↳which specific data visualization library or tool have you used the most? -
↳Selected Choice',
               'Approximately what percent of your time at work or school is
↳spent actively coding?',
               'How long have you been writing code to analyze data?',
               'For how many years have you used machine learning methods (at
↳work or in school)?',
               'Do you consider yourself to be a data scientist?',
               'What is the type of data that you currently interact with most
↳often at work or school? - Selected Choice',

```

```

        'How do you perceive the quality of online learning platforms_
↳and in-person bootcamps as compared to the quality of the education provided_
↳by traditional brick and mortar institutions? - In-person bootcamps:',
        'How do you perceive the importance of the following topics? -_
↳Fairness and bias in ML algorithms:',
        'How do you perceive the importance of the following topics? -_
↳Being able to explain ML model outputs and/or predictions',
        'How do you perceive the importance of the following topics? -_
↳Reproducibility in data science',
        'Approximately what percent of your data projects involve_
↳exploring model insights?',
        'Do you consider ML models to be "black boxes" with outputs that_
↳are difficult or impossible to explain?',
        'What is your age (# years)?',
        'In which country do you currently reside?',
        'What is the highest level of formal education that you have_
↳attained or plan to attain within the next 2 years?',
        'Which best describes your undergraduate major? - Selected_
↳Choice',
        'Select the title most similar to your current role (or most_
↳recent title if retired): - Selected Choice',
        'In what industry is your current employer/contract (or your_
↳most recent employer if retired)? - Selected Choice',
        'How many years of experience do you have in your current role?',
        'Does your current employer incorporate machine learning methods_
↳into their business?',
        'What is the primary tool that you use at work or school to_
↳analyze data? (include text response) - Selected Choice',
        'Approximately what percent of your data projects involved_
↳exploring unfair bias in the dataset and/or algorithm?',
        'Which better demonstrates expertise in data science: academic_
↳achievements or independent projects? - Your views:']

```

```

drop_cols_integer=['Duration (in seconds)',
        'What is the primary tool that you use at work or school to_
↳analyze data? (include text response) - Basic statistical software_
↳(Microsoft Excel, Google Sheets, etc.) - Text',
        'What is the primary tool that you use at work or school to_
↳analyze data? (include text response) - Advanced statistical software (SPSS,_
↳SAS, etc.) - Text',
        'What is the primary tool that you use at work or school to_
↳analyze data? (include text response) - Business intelligence software_
↳(Salesforce, Tableau, Spotfire, etc.) - Text',
        'What is the primary tool that you use at work or school to_
↳analyze data? (include text response) - Local or hosted development_
↳environments (RStudio, JupyterLab, etc.) - Text',

```

'What is the primary tool that you use at work or school to
 ↳analyze data? (include text response) - Cloud-based data software & APIs
 ↳(AWS, GCP, Azure, etc.) - Text',

'Which types of data do you currently interact with most
 ↳often at work or school? (Select all that apply) - Other Data - Text',

'What is the type of data that you currently interact with
 ↳most often at work or school? - Other Data - Text',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Gathering data',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Cleaning data',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Visualizing data',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Model building/model selection',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Putting the model into production',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Finding insights in the data and
 ↳communicating with stakeholders',

'During a typical data science project at work or school,
 ↳approximately what proportion of your time is devoted to the following?
 ↳(Answers must add up to 100%) - Other',

'What percentage of your current machine learning/data
 ↳science training falls under each category? (Answers must add up to 100%) -
 ↳Self-taught',

'What percentage of your current machine learning/data
 ↳science training falls under each category? (Answers must add up to 100%) -
 ↳Online courses (Coursera, Udemy, edX, etc.)',

'What percentage of your current machine learning/data
 ↳science training falls under each category? (Answers must add up to 100%) -
 ↳Work',

'What percentage of your current machine learning/data
 ↳science training falls under each category? (Answers must add up to 100%) -
 ↳University',

'What percentage of your current machine learning/data
 ↳science training falls under each category? (Answers must add up to 100%) -
 ↳Kaggle competitions',

```

        'What percentage of your current machine learning/data
        ↳science training falls under each category? (Answers must add up to 100%) -
        ↳Other']

target= ['What is your current yearly compensation (approximate $USD)?']
target_col=response_4.filter(target)

drop_cols_total=drop_cols_ohe+drop_cols_integer
label_encode_col=label_encode_col.drop(drop_cols_total, axis=1)

#### Label Encoding all the features that are possible
le=LabelEncoder()
label_encode_col=label_encode_col.apply(lambda col: le.fit_transform(col))

                                # ONE HOT ENCODING #

#### The following columns need to undergo one hot encoding
ohe_col=response_4.filter(drop_cols_ohe)
ohe_col_comp=pd.get_dummies(ohe_col,prefix=(drop_cols_ohe))
ohe_col_comp.shape

                                # INTEGER COLUMN CLEANING #

#Assigning a value of 0 or 1 where text was inputted but the text cannot be
↳seen
#When no text was inputted, -1 was assigned and when text was inputted, a
↳positive number was assigned
#Therefore, assigned 0 to all -1s (no text input) and 1 to all other numbers
↳(text was inputted)

integer_cols_comp=response_5.filter(drop_cols_integer)
numbers_col=['What is the primary tool that you use at work or school to
↳analyze data? (include text response) - Basic statistical software
↳(Microsoft Excel, Google Sheets, etc.) - Text',
            'What is the primary tool that you use at work or school to analyze
↳data? (include text response) - Advanced statistical software (SPSS, SAS,
↳etc.) - Text',
            'What is the primary tool that you use at work or school to analyze
↳data? (include text response) - Business intelligence software (Salesforce,
↳Tableau, Spotfire, etc.) - Text',
            'What is the primary tool that you use at work or school to analyze
↳data? (include text response) - Local or hosted development environments
↳(RStudio, JupyterLab, etc.) - Text',
            'What is the primary tool that you use at work or school to analyze
↳data? (include text response) - Cloud-based data software & APIs (AWS, GCP,
↳Azure, etc.) - Text',

```

```

        'Which types of data do you currently interact with most often at,
        ↳work or school? (Select all that apply) - Other Data - Text',
        'What is the type of data that you currently interact with most,
        ↳often at work or school? - Other Data - Text',
        'During a typical data science project at work or school,
        ↳approximately what proportion of your time is devoted to the following?
        ↳(Answers must add up to 100%) - Other']

text=integer_cols_comp.filter(numbers_col)
text=text.apply(pd.to_numeric)

#Changing all -1 (non-input) to 0 and all othe values to 1 (input):
text[text != -1] = 1
text[text == -1] = 0

#Normalizing the columns where the surveyer inputted number of hours and it,
↳adds upto 100
#Want all the numbers to be within 0 and 1 so it doesnt put more weight on the,
↳feature as compared to others.

#Normalize remaining integer columns and dropping duration (duration adds no,
↳value)
normalize_col=integer_cols_comp.drop(numbers_col,axis=1)
normalize_col=normalize_col.drop('Duration (in seconds)',axis=1)

#Removing null records because from observation it can be seen that records,
↳(wuth Naan) have naan records for all the integer column headers.
normalize_col.isnull().sum(axis=0)

#Normalizing the dataframe and changing it to int type
normalize_col=normalize_col.dropna()
normalize_col=normalize_col.apply(pd.to_numeric)
normalize_col = normalize_col.apply(lambda x: (x - x.min()) / (x.max() - x.
↳min()))
normalize_col.head(5)

#Joining both text and normalizing dataframes and only keeping records that are,
↳available in the normalize_col dataframe
integer_cols_comp=pd.merge(normalize_col, text,
↳left_index=True,right_index=True)

##### Joining all the datframes
#integer_cols_comp - has all the features that initially had integerts. The,
↳integers have been cleaned and normalzied
#ohe_col_comp - has all the features that underwent one hot encoding
#label_encode_col - has all the features that underwent label encoding

```



```
merged_v1=pd.merge(integer_cols_comp, ohe_col_comp,
    ↳left_index=True,right_index=True)
merged_v2=pd.merge(merged_v1,label_encode_col, left_index=True,right_index=True)
clean_data=pd.merge(merged_v2,target_col,left_index=True,right_index=True)

#Total number of records and features.
print ("\nTotal number of records: ", clean_data.shape[0])
print ("\nTotal number of features: ", clean_data.shape[1])
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/IPython/core/interactiveshell.py:3058: DtypeWarning: Columns (0,2,8,10,
21,23,24,25,26,27,28,44,56,64,83,85,87,107,109,123,125,150,157,172,174,194,210,2
18,219,223,246,249,262,264,276,277,278,279,280,281,282,283,284,285,286,287,288,2
89,290,304,306,325,326,329,341,368,371,384,385,389,390,391,393,394) have mixed
types. Specify dtype option on import or set low_memory=False.
interactivity=interactivity, compiler=compiler, result=result)
```

Total number of records: 11390

Total number of features: 664

2 EXPLORATORY DATA ANALYSIS

Four different graphical figures are created in order to understand the trend in the data. The following are the figures and the reasoning behind using it.

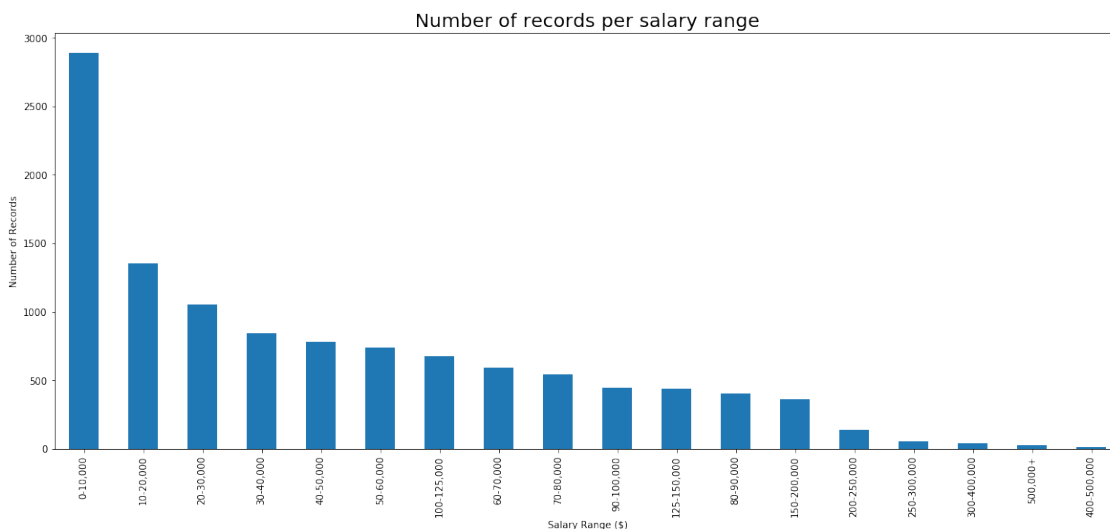
1. Number of records per salary range - This figure is very important as it helps us visualize the distribution of records per target variable. Upon analysis, I found that the distribution is severely imbalanced as there are more than 2900 records for "0-10,000" range and less than 50 records for 500,000+. Furthermore, it can be seen that the records become exponentially less once the salary range crosses 150,000. That is why I grouped together the salary range from 150,000 to 500,000+. This would balance the dataset and help the model predict accurate results. Another figure with the new distribution of records per salary range is also shown.
2. Number of records per country - This figure shows the number of records per country. In this case, I could not find any trend from this graph, other than the fact that US has too many records as compared to other countries. Since the requirement for each salary range could be different for different countries, I did not make any changes to the dataset.
3. Number of records per gender - This figure helped me determine that there was a very less percentage of records that did not claim any gender. As a result, those records were removed because they would not add any value. This helped me do label encoding on gender feature as there are only two responses (male and female). The updated distribution of gender ratio per record was plotted.
4. Degree distribution per records - This figure showed the distribution of degrees per records. Upon analysis I found that there were records that didn't prefer to answer. Those records

were removed in the data cleaning process.

These data visualization methods helped me understand the trend in the dataset and also clean it in more depth (after I was able to understand the relationships). Also, a correlation heatmap was plotted but since I have too many features, it became difficult to go through each correlations. As a result, I decided to use correlation in the feature selection Part.

```
[3]: #Bar plot to visualize the distribution of cleaned data
fig_1=clean_data['What is your current yearly compensation (approximate $USD)?
↪_y'].value_counts().plot(kind='bar',figsize=(20,8))
plt.xlabel('Salary Range ($)')
plt.ylabel('Number of Records')
plt.title('Number of records per salary range',fontsize=20)
```

```
[3]: Text(0.5, 1.0, 'Number of records per salary range')
```



As there seems to be an uneven distribution of data, it can be seen that the records become exponentially less once the salary range crosses \$150,000. That is why I grouped together the salary range from 200,000 to 500,000+. This would help me keep majority of the samples in the rare class (salary ranges with less records) while randomly selecting an equal number of samples from the abundant class. A balanced new dataset can be retrieved for further modelling. This would yield accurate results.

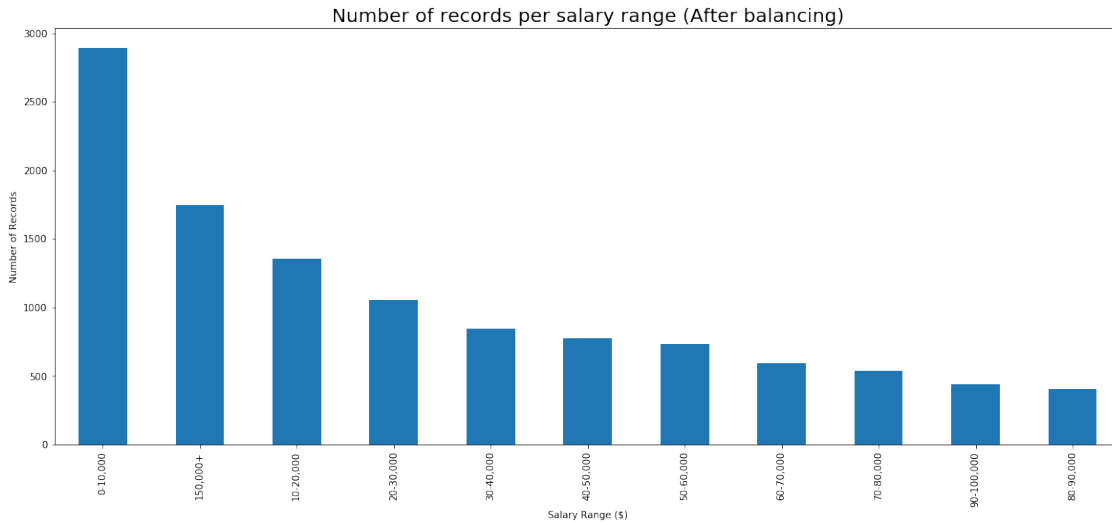
```
[4]: #Grouping salary buckets of 200,000 to 500,000
clean_data['What is your current yearly compensation (approximate $USD)?
↪_y']=clean_data['What is your current yearly compensation (approximate $USD)?
↪_y'].replace({'100-125,000':'150,000+', '125-150,000':
↪'150,000+', '150-200,000':'150,000+', '200-250,000':'150,000+', '250-300,000':
↪'150,000+', '300-400,000':'150,000+', '400-500,000':'150,000+', '500,000+':
↪'150,000+' })
```

```

#Plot representing new distribution of data:
#Bar plot to visualize the distribution of cleaned data
fig_2=clean_data['What is your current yearly compensation (approximate $USD)?
→_y'].value_counts().plot(kind='bar',figsize=(20,8))
plt.xlabel('Salary Range ($)')
plt.ylabel('Number of Records',fontsize=10)
plt.title('Number of records per salary range (After balancing)',fontsize=20)

```

[4]: Text(0.5, 1.0, 'Number of records per salary range (After balancing)')

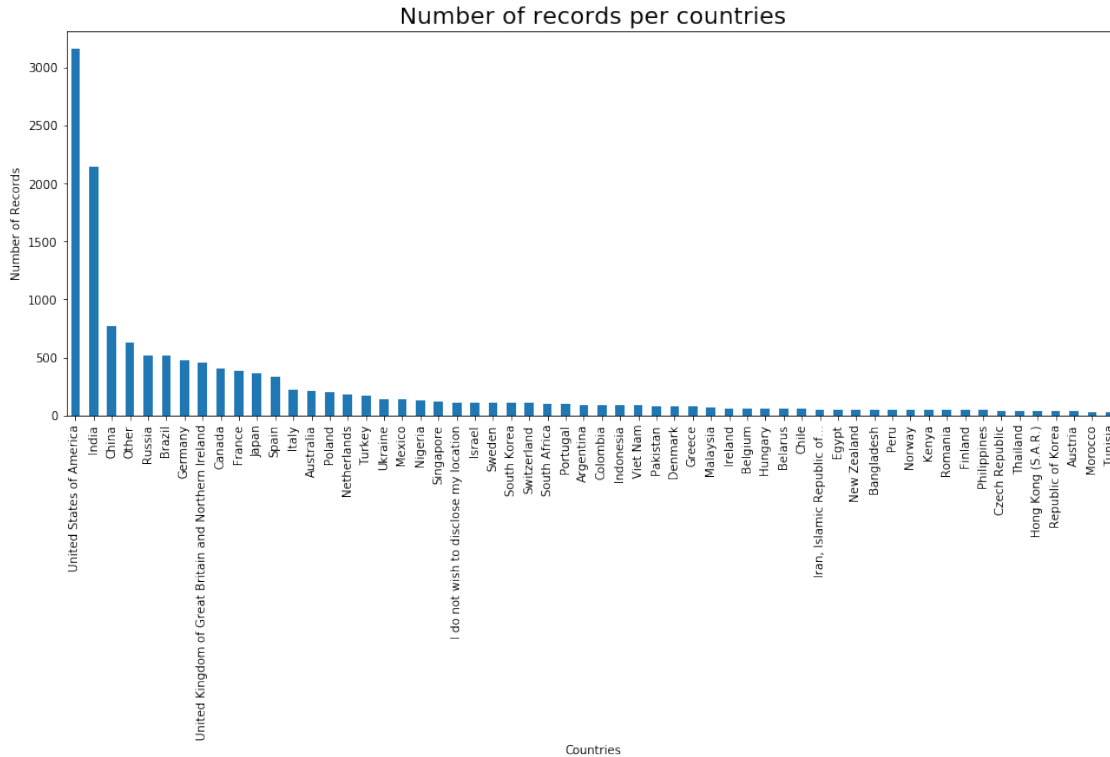


```

[5]: #Bar plot to visualize the distribution of cleaned data
fig_1=response_5['In which country do you currently reside?'].value_counts().
→plot(kind='bar',figsize=(16,6))
plt.xlabel('Countries',fontsize=10)
plt.ylabel('Number of Records',fontsize=10)
plt.title('Number of records per countries',fontsize=20)

```

[5]: Text(0.5, 1.0, 'Number of records per countries')



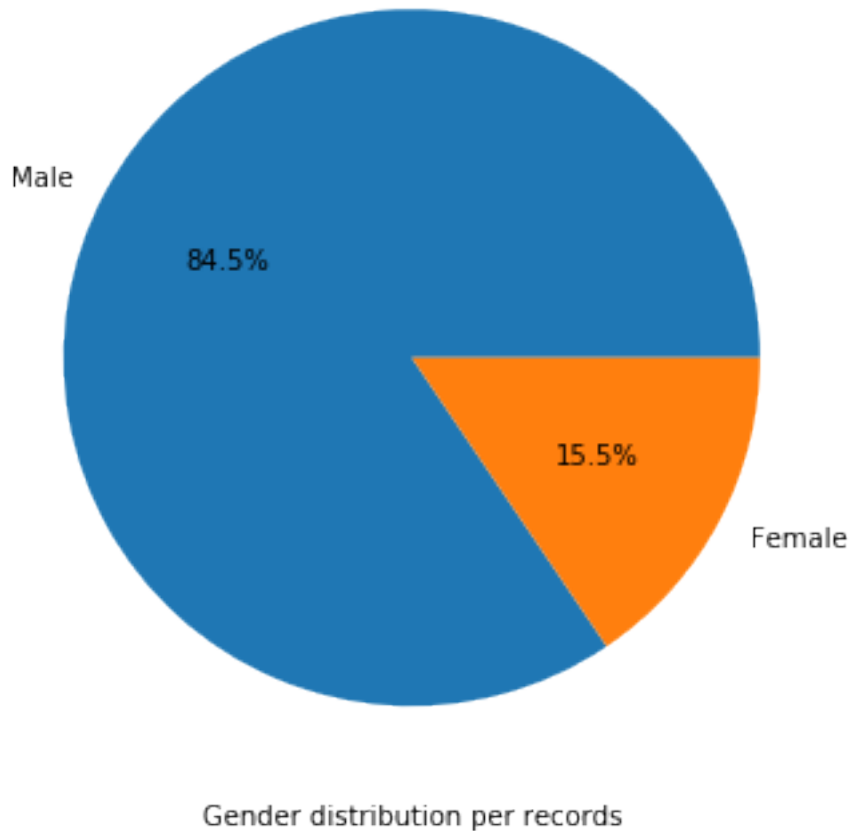
From the above graph, it can be seen that United States of America has a lot of records in comparison to other countries. The distribution is severely uneven but I will not use any technique as it does not concern the target variable (salary).

Next, I will try to see the gender distribution distribution per records. I initially found that 0.8% of the records were either "Prefer to self-describe" and "Prefer not to say". As a results these records were initially removed in Data Cleaning Part II. Resulting pie chart can be see below. The removed records would not add any value and their removal would yield better results for our model.

```
[6]: #Bar plot to visualize the distribution of cleaned data
fig_1=response_5['What is your gender? - Selected Choice'].value_counts().
    ↳plot(kind='pie',autopct='%1.1f%%',figsize=(16,6))
plt.xlabel('Gender distribution per records',fontsize=10)
plt.title('Number of records per gender',fontsize=20)
plt.ylabel('',fontsize=10)
```

```
[6]: Text(0, 0.5, '')
```

Number of records per gender



Next, I will try to see the degree distribution per records. I initially found that a very low percentage of records were "I prefer not to answer". As a result, these records were removed in Data Cleaning Part II because they would not add any value to our model. Resulting bar chart can be seen below.

```
[7]: #Bar plot to visualize the distribution of cleaned data
fig_1=response_5['What is the highest level of formal education that you have_
→attained or plan to attain within the next 2 years?'].value_counts().
→plot(kind='bar',figsize=(16,6))
plt.xlabel('Degree Distribution',fontsize=10)
plt.ylabel('Number of Records',fontsize=10)
plt.title('Degree distribution per records',fontsize=20)
```

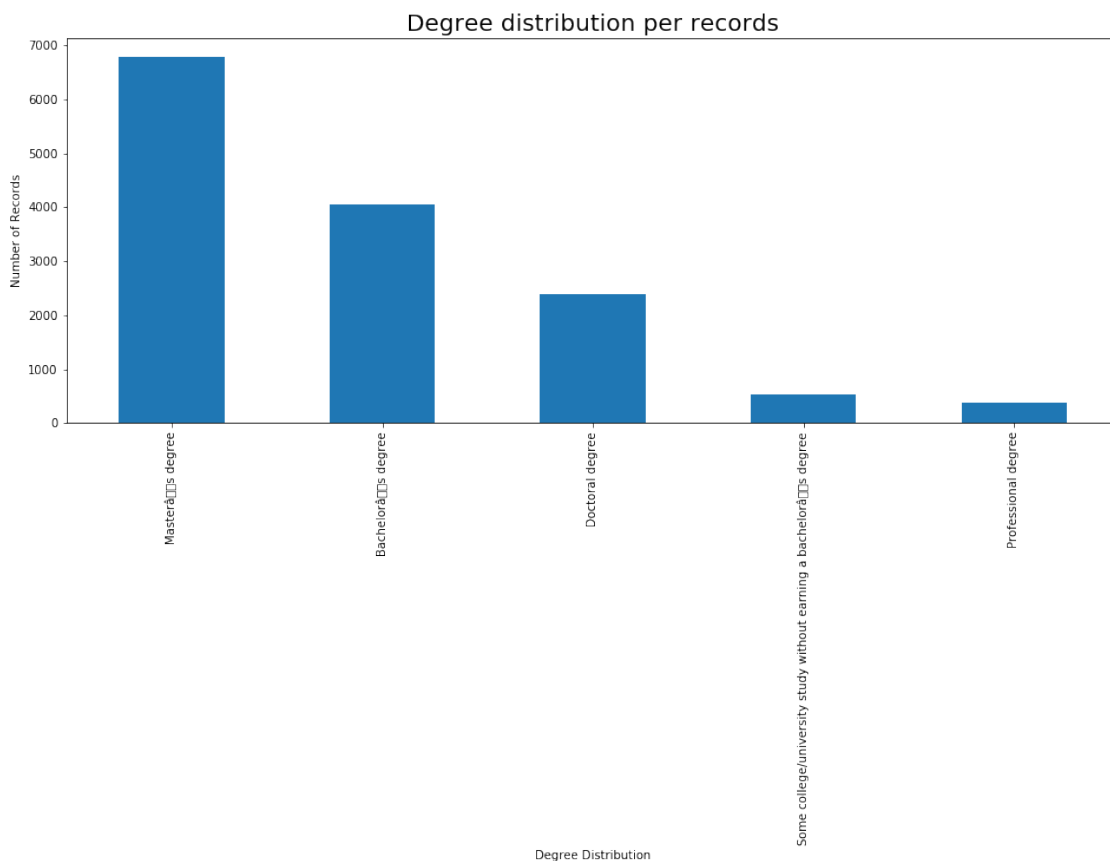
```
[7]: Text(0.5, 1.0, 'Degree distribution per records')
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/matplotlib/backends/backend_agg.py:211: RuntimeWarning: Glyph 128
missing from current font.
```

```

font.set_text(s, 0.0, flags=flags)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/matplotlib/backends/backend_agg.py:211: RuntimeWarning: Glyph 153
missing from current font.
font.set_text(s, 0.0, flags=flags)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/matplotlib/backends/backend_agg.py:180: RuntimeWarning: Glyph 128
missing from current font.
font.set_text(s, 0, flags=flags)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/matplotlib/backends/backend_agg.py:180: RuntimeWarning: Glyph 153
missing from current font.
font.set_text(s, 0, flags=flags)

```



In order to visualize the importance of each feature, a correlation plot was plotted as a heat map. The correlation plot is too big as there are too many features. That is why I commented it out as it will become very slow if correlation plot is plotted. The correlation plot will be used in the feature selection section.

```

[8]: #Correlation matrix
      #corr=clean_data.corr().abs()

```

```
#Creating correleation matrix  
#corr.style.background_gradient(cmap='coolwarm',low=0).set_precision(2)
```

3 FEATURE SELECTION

Feature engineering is a very useful tool in machine learning. It is a process of extracting the most relevant features from the dataset and then applying machine learning algorithms for the better performance of the model. A large number of irrelevant features increases the training time exponentially and will also increase the risk of overfitting.

I did feature engineering in multiple steps. Initially, a correlation matrix was plotted and then features with correlation coefficient of more than 70% was identified. After that, one of them was removed (from the pair). Removing features with high colinearity will improve the performance of our model as keeping both would be counter-productive.

After reducing the number of features bases on high colinearity, I used Chi Square test for feature extraction. Chi-square test is used for categorical features in a dataset. In my case, the target is categorical (I kept it as it is). We calculate Chi-square between each feature and the target and select the desired number of features with best Chi-square scores. It determines if the association between two categorical variables of the sample would reflect their real association in the population.

After identifying desired number of featur, Principal Component Analysis was used to reduce the dimensionality of data. It would help in proving a better estimate. I would like to clarify that I used Chi square test to first identify important features from the dataset and then I used PCA to reduce the dimensionality of the dataset wihtout losing any information.

```
[9]: #Correlation matrix  
#Removing features with high colinearity using correleation matrix  
corr=clean_data.corr().abs()  
  
#Creating correleation matrix  
corr.style.background_gradient(cmap='coolwarm',low=0).set_precision(2)  
  
#Select upper triangle of correleation matrix  
upper = corr.where(np.triu(np.ones(corr.shape), k=1).astype(np.bool))  
  
#Finding index of feature columns with correlation greater than 0.70  
to_drop = [column for column in upper.columns if any(upper[column] > 0.7)]  
  
#Dropping identified features  
clean_data_v1=clean_data.drop(clean_data[to_drop], axis=1)  
  
#Chi-Squared for Feature selection  
# Create features and target  
target_sal=['What is your current yearly compensation (approximate $USD)?_y']  
y=clean_data_v1.filter(target_sal)  
x=clean_data_v1.drop(target_sal,axis=1)
```

```

#Selecting features with highest chi-squared statistics (selected number is 50)
x_new = SelectKBest(chi2, k=10).fit_transform(x, y)
x_chi=pd.DataFrame(x_new)

#Principal component analysis
from sklearn.decomposition import PCA
pca = PCA(n_components=5)
x = pca.fit_transform(x_chi)

#Print number of features:
print ("\nTotal number of features after dropping features with high_
↪colinearity: ", clean_data_v1.shape[1])
print ("\nTotal number of features after using Chi Square test ", x_chi.
↪shape[1])
print ("\nTotal number of features after Principal Component Analysis ", x.
↪shape[1])

```

Total number of features after dropping features with high colinearity: 648

Total number of features after using Chi Square test 10

Total number of features after Principal Component Analysis 5

4 MODEL IMPLEMENTATION

Logistic Regression model is used as specified in the problem set. As there is an uneven distribution of samples, `class_weight="balanced"` is used as it basically means replicating the smaller class until you have as many samples as the larger one. Regularization is also used as it is a way to avoid overfitting by penalizing high-valued regression coefficients. It reduces parameters and simplifies the model. Regularization is necessary because least squares regression methods, where the residual sum of squares is minimized, can be unstable. This is especially true if there is multicollinearity in the model. In this case, Lasso regression is used. L1 regularization adds an L1 penalty equal to the absolute value of the magnitude of coefficients. In other words, it limits the size of the coefficients. All the features are normalized in the preprocessing part, therefore; I do not need to scale/normalize anything here.

Lastly, 10 fold stratified cross validation is used and precision scores across all the folds are printed along with average accuracy and variance of accuracy.

```

[10]: #Split x and y into training and testing sets
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)

#Create and fit a model
logmodel = LogisticRegression(C=1,penalty='l1',class_weight =_
↪"balanced",solver='liblinear',multi_class='ovr',max_iter=200)

```



```

logmodel.fit(X_train, y_train)

#Cross validation (10 fold)
skf = StratifiedKFold(n_splits=10)
accuracy = cross_val_score(logmodel, x, y, cv=skf, scoring='accuracy')
    ↳#Accuracy across folds
avg_accuracy= cross_val_score(logmodel, x, y, cv=skf,scoring='accuracy').mean()
    ↳#Average accuracy for folds
avg_variance= cross_val_score(logmodel, x, y, cv=skf,scoring='accuracy').var()
    ↳#Average variance of accuracy
avg_recall = cross_val_score(logmodel, x, y, cv=skf, scoring='recall_macro',).
    ↳mean() #Average recall for folds
avg_precision = cross_val_score(logmodel, x, y, cv=skf,
    ↳scoring='precision_macro').mean() #Average precision for folds
avg_f1 = cross_val_score(logmodel, x, y, cv=skf, scoring='f1_macro').mean()
    ↳#Average f1 for folds

#Predictions (Probability of belonging to each salary bucket is printed)
Pred_cross_val = cross_val_predict(logmodel, x, y, cv=skf)

print ('\nResults:')
print ('\nAccuracy across each fold: ', accuracy)
print ('\nAverage accuracy across 10 folds: ', round(avg_accuracy,2))
print ('\nVariance of accuracy across 10 folds: ', round(avg_variance,5))
print ('\nAverage recall across 10 folds: ', round(avg_recall,2))
print ('\nAverage precision across 10 folds: ', round(avg_precision,2))
print ('\nAverage F1 score across 10 folds: ', round(avg_f1,2))
print ('\nProbability of belonging to each salary bucket:')
print(classification_report(y,Pred_cross_val))

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.

```
"the number of iterations.", ConvergenceWarning)
```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.

```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```



```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```



```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```

```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)

```

Results:

```

Accuracy across each fold: [0.81222707 0.80157343 0.75153106 0.80122592
0.81666667 0.76382792]

```

```
0.78502203 0.79365079 0.77336861 0.84567901]
```

Average accuracy across 10 folds: 0.8

Variance of accuracy across 10 folds: 0.00069

Average recall across 10 folds: 0.73

Average precision across 10 folds: 0.74

Average F1 score across 10 folds: 0.73

Probability of belonging to each salary bucket:

	precision	recall	f1-score	support
0-10,000	1.00	1.00	1.00	2894
10-20,000	0.80	0.90	0.85	1356
150,000+	0.78	0.71	0.74	1746
20-30,000	0.80	0.77	0.79	1056
30-40,000	0.71	0.66	0.69	844
40-50,000	0.70	0.73	0.72	777
50-60,000	0.63	0.79	0.70	735
60-70,000	0.35	0.36	0.35	592
70-80,000	0.42	0.37	0.39	541
80-90,000	0.95	0.74	0.83	406
90-100,000	1.00	1.00	1.00	443
micro avg	0.79	0.79	0.79	11390
macro avg	0.74	0.73	0.73	11390
weighted avg	0.80	0.79	0.79	11390

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

In the section below, I will be manually tuning the hyperparameter (C) for logistic regression to find the model that performed the best. The selected model will be justified using bias-variance tradeoff.

```
[11]: #Hyperparameter tuning:  
#Changing values of C  
logmodel_c1 = LogisticRegression(C=1,penalty='l1',class_weight =_  
    ↪ "balanced",solver='liblinear',multi_class='ovr',max_iter=200)  
logmodel_c2 = LogisticRegression(C=5,penalty='l1',class_weight =_  
    ↪ "balanced",solver='liblinear',multi_class='ovr',max_iter=200)
```

```

logmodel_c3 = LogisticRegression(C=10,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c4 = LogisticRegression(C=20,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c5 = LogisticRegression(C=40,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c6 = LogisticRegression(C=80,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c7 = LogisticRegression(C=100,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c8 = LogisticRegression(C=150,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c9 = LogisticRegression(C=200,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c10 = LogisticRegression(C=250,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c15 = LogisticRegression(C=300,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c20 = LogisticRegression(C=500,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c30 = LogisticRegression(C=1000,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c40 = LogisticRegression(C=2000,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel_c50 = LogisticRegression(C=5000,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)

#New cross validated accuracies:
accuracy_C1 = cross_val_score(logmodel_c1, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C2 = cross_val_score(logmodel_c2, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C3 = cross_val_score(logmodel_c3, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C4 = cross_val_score(logmodel_c4, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C5 = cross_val_score(logmodel_c5, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C6 = cross_val_score(logmodel_c6, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C7 = cross_val_score(logmodel_c7, x, y, cv=skf, scoring='accuracy').
    ↳mean()
accuracy_C8 = cross_val_score(logmodel_c8, x, y, cv=skf, scoring='accuracy').
    ↳mean()

```



```

accuracy_C9 = cross_val_score(logmodel_c9, x, y, cv=skf, scoring='accuracy').
↳mean()
accuracy_C10 = cross_val_score(logmodel_c10, x, y, cv=skf, scoring='accuracy').
↳mean()
accuracy_C15 = cross_val_score(logmodel_c15, x, y, cv=skf, scoring='accuracy').
↳mean()
accuracy_C20 = cross_val_score(logmodel_c20, x, y, cv=skf, scoring='accuracy').
↳mean()
accuracy_C30 = cross_val_score(logmodel_c30, x, y, cv=skf, scoring='accuracy').
↳mean()
accuracy_C40 = cross_val_score(logmodel_c40, x, y, cv=skf, scoring='accuracy').
↳mean()
accuracy_C50 = cross_val_score(logmodel_c50, x, y, cv=skf, scoring='accuracy').
↳mean()

#Printing accuracy:
print ('\nAccuracies of different models when hyperparameter is tuned:')
print ('\nAverage accuracy when C=1:', round(accuracy_C1,5))
print ('\nAverage accuracy when C=5:', round(accuracy_C2,5))
print ('\nAverage accuracy when C=10:', round(accuracy_C3,5))
print ('\nAverage accuracy when C=20:', round(accuracy_C4,5))
print ('\nAverage accuracy when C=40:', round(accuracy_C5,5))
print ('\nAverage accuracy when C=80:', round(accuracy_C6,5))
print ('\nAverage accuracy when C=100:', round(accuracy_C7,5))
print ('\nAverage accuracy when C=150:', round(accuracy_C8,5))
print ('\nAverage accuracy when C=200:', round(accuracy_C9,5))
print ('\nAverage accuracy when C=250:', round(accuracy_C10,5))
print ('\nAverage accuracy when C=300:', round(accuracy_C15,5))
print ('\nAverage accuracy when C=500:', round(accuracy_C20,5))
print ('\nAverage accuracy when C=1000:', round(accuracy_C30,5))
print ('\nAverage accuracy when C=2000:', round(accuracy_C40,5))
print ('\nAverage accuracy when C=5000:', round(accuracy_C50,5))

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.

```
"the number of iterations.", ConvergenceWarning)
```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```

```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```

```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```

```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```



```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)

```

Accuracies of different models when hyperparameter is tuned:

Average accuracy when C=1: 0.79527

Average accuracy when C=5: 0.79125

Average accuracy when C=10: 0.79134

Average accuracy when C=20: 0.79134

Average accuracy when C=40: 0.79142

Average accuracy when C=80: 0.79142

Average accuracy when C=100: 0.79142

Average accuracy when C=150: 0.79142

Average accuracy when C=200: 0.79142

Average accuracy when C=250: 0.79142

Average accuracy when C=300: 0.79125

Average accuracy when C=500: 0.79142

Average accuracy when C=1000: 0.79142

Average accuracy when C=2000: 0.79125

Average accuracy when C=5000: 0.79142

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.

"the number of iterations.", ConvergenceWarning)

The best model is when the value of hyperparameter (C) is 1. Below, learning curve using this parameter is plotted in order to show bias-variance. Both the training and validation curves are close to each other and converges at low complexity. This shows that the model is working well and is not underfitting or overfitting.

```
[12]: import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import learning_curve
from sklearn.model_selection import ShuffleSplit

def plot_learning_curve(estimator, title, X, y, ylim=None, cv=None,
                        n_jobs=None, train_sizes=np.linspace(.1, 1.0, 5)):
    plt.figure()
    plt.title(title)
    if ylim is not None:
        plt.ylim(*ylim)
    plt.xlabel("Training examples")
    plt.ylabel("Score")
    train_sizes, train_scores, test_scores = learning_curve(
        estimator, X, y, cv=cv, n_jobs=n_jobs, train_sizes=train_sizes)
```

```

train_scores_mean = np.mean(train_scores, axis=1)
train_scores_std = np.std(train_scores, axis=1)
test_scores_mean = np.mean(test_scores, axis=1)
test_scores_std = np.std(test_scores, axis=1)
plt.grid()
plt.fill_between(train_sizes, train_scores_mean - train_scores_std,
                 train_scores_mean + train_scores_std, alpha=0.1,
                 color="r")
plt.fill_between(train_sizes, test_scores_mean - test_scores_std,
                 test_scores_mean + test_scores_std, alpha=0.1, color="g")
plt.plot(train_sizes, train_scores_mean, 'o-', color="r",
         label="Training score")
plt.plot(train_sizes, test_scores_mean, 'o-', color="g",
         label="Cross-validation score")
plt.legend(loc="best")
return plt

# Cross validation with 100 iterations to get smoother mean test and train
# score curves, each time with 20% data randomly selected as a validation set.
cv = ShuffleSplit(n_splits=100, test_size=0.2, random_state=0)

estimator = LogisticRegression(C=1,penalty='l1',class_weight =_
    ↪"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
plot_learning_curve(estimator, 'Logistic Regression Learning Curve', x, y,_
    ↪ylim=(0, 1), cv=10, n_jobs=1)

from matplotlib.pyplot import figure
plt.figure()

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().

```

```

y = column_or_1d(y, warn=True)

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.

```

```

"the number of iterations.", ConvergenceWarning)

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().

```

```

y = column_or_1d(y, warn=True)

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.

```

```

"the number of iterations.", ConvergenceWarning)

```



```

(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-

```

```

packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)

```

```

(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-

```

```

packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)

```



```

(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-

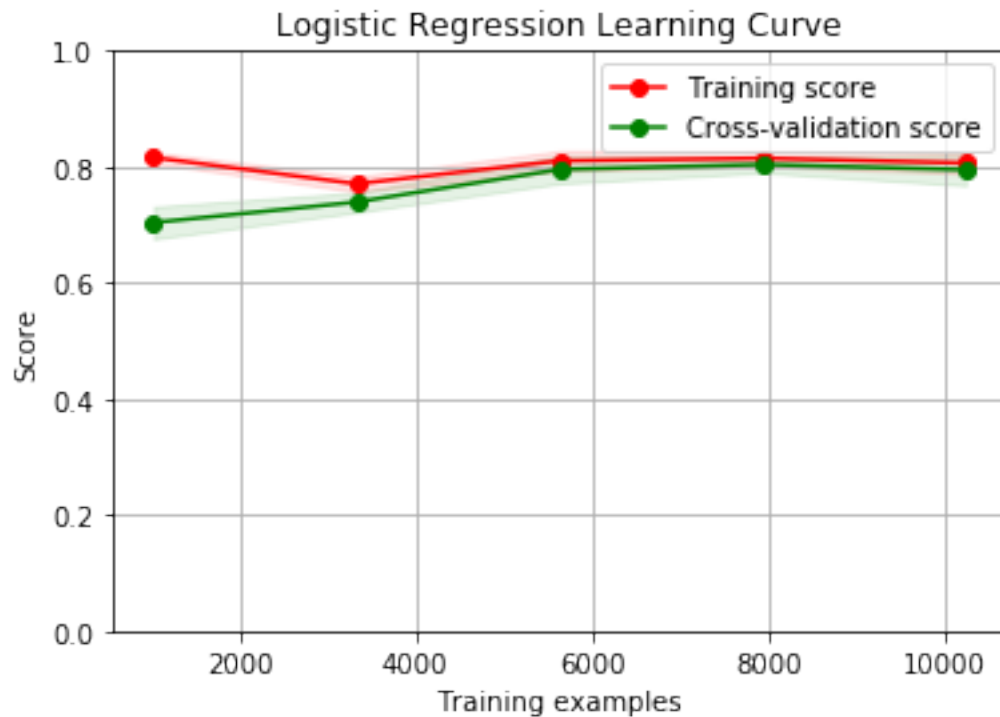
```

```

packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)

```


[12]: <Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>

5 MODEL TUNING

In order to tune the model, a grid search is used to determine the optimal model selected. The accuracy metrics selected are precision and recall. I would want to correctly identify the salary range based on the skillset of the candidate. Also, the grid search is commented out and the results are printed below. The hyperparameters that are being tuned are C, penalty and class_weight.

```
[13]: # Grid search cross validation
# from sklearn.model_selection import GridSearchCV
# from sklearn.linear_model import LogisticRegression

# grid={"C":np.logspace(-10,10), "penalty":["l1","l2"], "class_weight":[None,
# ↪ "balanced"]}
# logreg= LogisticRegression(solver='liblinear',multi_class='ovr',max_iter=200)
# logreg_cv=GridSearchCV(logreg,grid,cv=skf)
# logreg_cv.fit(X_train,y_train)
```

```
#logmodel = LogisticRegression(C=50, penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr'
logreg_cv.fit(X_train,y_train)

#print("Tuned hpyerparameters :(best parameters) ",logreg_cv.best_params_)
#print("Best Accuracy :",logreg_cv.best_score_)
```

Based on grid search, the following are the results: Tuned hpyerparameters :(best parameters) {'C': 0.03727593720314938, 'class_weight': 'balanced', 'penalty': 'l1'} Best Accuracy : 0.8594004766085539

Using tuned hyper-parameters, the following are the crossvalidated scores of the metrics chosen to select the model.

```
[14]: logmodel = LogisticRegression(C=0.04,penalty='l1',class_weight =_
    ↳"balanced",solver='liblinear',multi_class='ovr',max_iter=200)
logmodel.fit(X_train, y_train)
avg_accuracy= cross_val_score(logmodel, x, y, cv=skf,scoring='accuracy').mean()_
    ↳#Average accuracy for folds
avg_recall = cross_val_score(logmodel, x, y, cv=skf, scoring='recall_macro',).
    ↳mean() #Average recall for folds
avg_precision = cross_val_score(logmodel, x, y, cv=skf,_
    ↳scoring='precision_macro').mean() #Average precision for folds
avg_f1 = cross_val_score(logmodel, x, y, cv=skf, scoring='f1_macro').mean()_
    ↳#Average f1 for folds

print ('Average accuracy across 10 folds:', round(avg_accuracy,2))
print ('Average recall across 10 folds: ', round(avg_recall,2))
print ('Average precision across 10 folds: ', round(avg_precision,2))
print ('Average F1 score across 10 folds: ', round(avg_f1,2))
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
```



```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.  
    "the number of iterations.", ConvergenceWarning)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector  
y was passed when a 1d array was expected. Please change the shape of y to  
(n_samples, ), for example using ravel().  
    y = column_or_1d(y, warn=True)  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-  
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to  
converge, increase the number of iterations.
```



```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

Average accuracy across 10 folds: 0.85
Average recall across 10 folds: 0.84
Average precision across 10 folds: 0.85
Average F1 score across 10 folds: 0.84

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)

```

6 Testing & Discussion

The optimal model that have been determined from the previous sections will be used to make classification on the test set. The accuracy of the model on both the train and test set will is shown as well along with the learning curve.

```

[15]: #Accuracy on training set
print ('Accuracy on training set: (To see if the model is overfitting or
      ↳underfitting):')
train_pred = logmodel.predict(X_train)
print(classification_report(y_train,train_pred))
print('\nSince the accuracy of the training set is pretty close to the accuracy
      ↳of the testing set, it can be safely concluded that the model is not
      ↳underfitting or overfitting.')

```

```

Accuracy on training set: (To see if the model is overfitting or underfitting):
      precision    recall  f1-score   support

```

0-10,000	0.91	0.99	0.95	2037
10-20,000	0.78	0.74	0.76	957
150,000+	0.77	0.71	0.74	1218
20-30,000	0.86	0.83	0.85	727
30-40,000	0.85	0.80	0.83	618
40-50,000	0.83	0.81	0.82	517
50-60,000	0.84	0.84	0.84	487

60-70,000	0.80	0.61	0.69	418
70-80,000	0.77	0.98	0.86	386
80-90,000	1.00	1.00	1.00	291
90-100,000	1.00	1.00	1.00	317
micro avg	0.85	0.85	0.85	7973
macro avg	0.85	0.85	0.85	7973
weighted avg	0.85	0.85	0.85	7973

Since the accuracy of the training set is pretty close to the accuracy of the testing set, it can be safely concluded that the model is not underfitting or overfitting.

```
[16]: def plot_learning_curve(estimator, title, X, y, ylim=None, cv=None,
                             n_jobs=None, train_sizes=np.linspace(.1, 1.0, 5)):
    plt.figure()
    plt.title(title)
    if ylim is not None:
        plt.ylim(*ylim)
    plt.xlabel("Training examples")
    plt.ylabel("Score")
    train_sizes, train_scores, test_scores = learning_curve(
        estimator, X, y, cv=cv, n_jobs=n_jobs, train_sizes=train_sizes)
    train_scores_mean = np.mean(train_scores, axis=1)
    train_scores_std = np.std(train_scores, axis=1)
    test_scores_mean = np.mean(test_scores, axis=1)
    test_scores_std = np.std(test_scores, axis=1)
    plt.grid()
    plt.fill_between(train_sizes, train_scores_mean - train_scores_std,
                     train_scores_mean + train_scores_std, alpha=0.1,
                     color="r")
    plt.fill_between(train_sizes, test_scores_mean - test_scores_std,
                     test_scores_mean + test_scores_std, alpha=0.1, color="g")
    plt.plot(train_sizes, train_scores_mean, 'o-', color="r",
             label="Training score")
    plt.plot(train_sizes, test_scores_mean, 'o-', color="g",
             label="Cross-validation score")
    plt.legend(loc="best")
    return plt

cv = skf
estimator = LogisticRegression(C=0.04,penalty='l1',class_weight =_
    ↪ "balanced",solver='liblinear',multi_class='ovr',max_iter=200)
plot_learning_curve(estimator, 'Logistic Regression Learning Curve', x, y,_
    ↪ ylim=(0, 1), cv=skf, n_jobs=1)
```

```
from matplotlib.pyplot import figure
plt.figure()
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
```

```

packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().

```

```

y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)

```

```

/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-

```



```

packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector

```

```

y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.

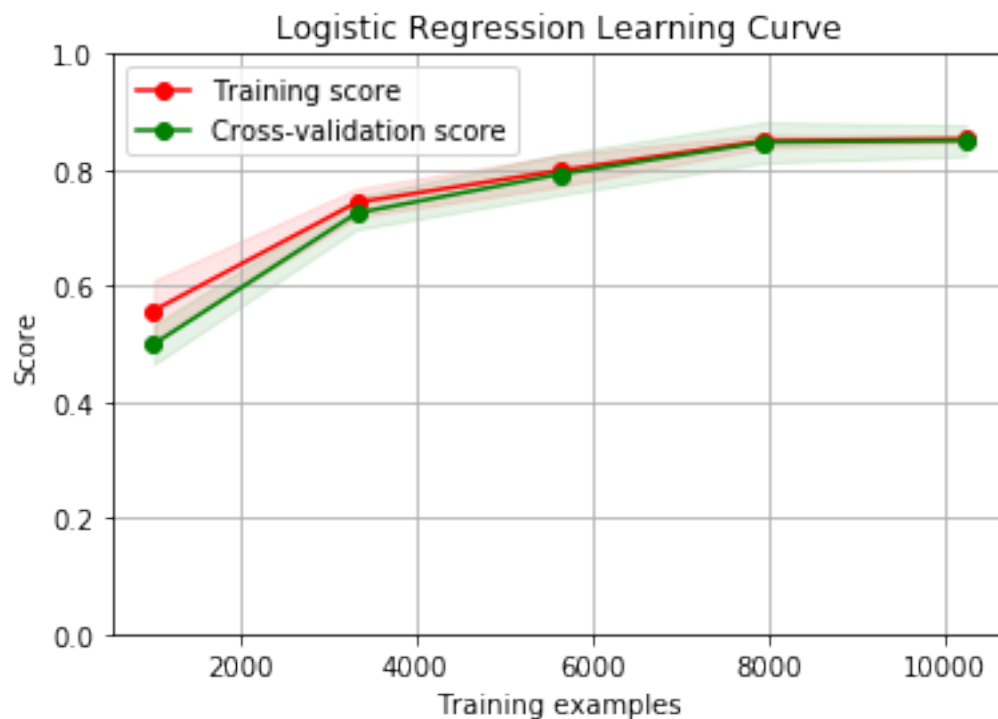
```

```

    "the number of iterations.", ConvergenceWarning)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/utils/validation.py:761: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/sklearn/svm/base.py:922: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)

```

[16]: <Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>

After completing all the steps, I can confidently conclude that the overall fit of the model is good. It is not overfitting or underfitting as can be seen from the learning curve above. The model also converges to an acceptable accuracy score at low complexity. Initially, I was getting accuracy of 70 percent. After tuning the hyperparameters, I am getting an accuracy of 85 percent with the help of grid search.

Furthermore, the accuracy of the model could be increased by balancing the dataset using under-sampling techniques such as SMOTE or ADASYN. Synthetic minority oversampling technique or Adaptive synthetic sampling method would create more samples for the minority class by taking

existing information from the existing samples. A balanced dataset would provide higher unbiased accuracy. Lastly, there is a lot of noise present in the dataset that should be removed. Refining the dataset would increase accuracy. Examples of noise would include records of students having a salary of more than 250,000+. Those records should be moved. Also, I personally feel that the dataset is misleading. Salary range is geographic as someone with the same skillset could be earning a lot more in Western countries than Eastern countries. As a result, the model would get confused and give inaccurate results. That is why the model should be implemented individually for a specific geographic location in order to get accurate estimates.

[]: