Dependency Package

In [1]:
```
import pandas as pd
import numpy as np
%matplotlib inline

#!pip install lime
import lime.lime_tabular

from scipy.stats import chi2_contingency, kruskal, f_oneway, normaltest, bartlett
import plotly.express as px
import plotly.graph_objects as go

import sys
#!{sys.executable} -m pip install keras
#!{sys.executable} -m pip install -U keras-tuner
import kerastuner
from kerastuner import RandomSearch

import tensorflow as tf
from tensorflow import keras
from keras import layers
from sklearn.model_selection import StratifiedKFold

#!{sys.executable} -m pip install plotly

from sklearn.tree import DecisionTreeClassifier, export_graphviz
from six import StringIO
from IPython.display import Image
import pydotplus

tf.random.set_seed(2)
```

Explore Data

In [2]:
```
FILE1 = 'assessments.csv'
#FILE2 = 'courses.csv'                #irrelevant
FILE3 = 'studentAssessment.csv'
FILE4 = 'studentInfo.csv'
#FILE5 = 'studentRegistration.csv'    #irrelevant
FILE6 = 'studentVle.csv'
#FILE7 = 'vle.csv'                    #irrelevant
```

In [3]:
```
assessment_df = pd.read_csv(FILE1)
#courses_df = pd.read_csv(FILE2)
studentAssessment_df = pd.read_csv(FILE3)
studentInfo_df = pd.read_csv(FILE4)
#studentRegistration_df = pd.read_csv(FILE5)
studentVle_df = pd.read_csv(FILE6)
#vle_df = pd.read_csv(FILE7)
```

In [4]:
```
display(assessment_df.head())
print(assessment_df.shape)
```

```
assessment_df.isnull().sum()
```

	code_module	code_presentation	id_assessment	assessment_type	date	weight
0	AAA	2013J	1752	TMA	19	10.0
1	AAA	2013J	1753	TMA	54	20.0
2	AAA	2013J	1754	TMA	117	20.0
3	AAA	2014J	1758	TMA	19	10.0
4	AAA	2014J	1759	TMA	54	20.0

Out[4]:
```
(37, 6)
code_module          0
code_presentation    0
id_assessment        0
assessment_type      0
date                 0
weight               0
dtype: int64
```

In [5]:
```
display(studentAssessment_df.head())
studentAssessment_df.shape
```

	id_assessment	id_student	date_submitted	is_banked	score
0	1752	11391	18	0	78.0
1	1752	28400	22	0	70.0
2	1752	31604	17	0	72.0
3	1752	32885	26	0	69.0
4	1752	38053	19	0	79.0

Out[5]:
```
(36184, 5)
```

In [6]:
```
studentAssessment_df.isnull().sum()
```

Out[6]:
```
id_assessment      0
id_student         0
date_submitted     0
is_banked          0
score             18
dtype: int64
```

In [7]:
```
display(studentInfo_df.head())
print(studentInfo_df.shape)
studentInfo_df.isnull().sum()
```

	code_module	code_presentation	id_student	gender	region	highest_education	imd_band	age_
0	AAA	2013J	11391	M	East Anglian Region	HE Qualification	90-100%	
1	AAA	2013J	28400	F	Scotland	HE Qualification	20-30%	
2	AAA	2013J	31604	F	South East Region	A Level or Equivalent	50-60%	
3	AAA	2013J	32885	F	West Midlands Region	Lower Than A Level	50-60%	
4	AAA	2013J	38053	M	Wales	A Level or Equivalent	80-90%	

```
(12489, 12)
```

Out[7]:
```
code_module              0
code_presentation        0
id_student               0
gender                   0
region                   0
highest_education        0
imd_band               422
age_band                 0
num_of_prev_attempts     0
studied_credits          0
disability               0
final_result             0
dtype: int64
```

In [8]:
```
display(studentVle_df.head())
print(studentVle_df.shape)
studentVle_df.isnull().sum()
```

	code_module	code_presentation	id_student	id_site	date	sum_click
0	AAA	2013J	28400	546652	-10	4
1	AAA	2013J	28400	546652	-10	1
2	AAA	2013J	28400	546652	-10	1
3	AAA	2013J	28400	546614	-10	11
4	AAA	2013J	28400	546714	-10	1

```
(3315787, 6)
```

Out[8]:
```
code_module          0
code_presentation    0
id_student           0
id_site              0
date                 0
sum_click            0
dtype: int64
```

In [9]:
```
studentVle_df = studentVle_df.drop(columns=['id_site', 'date'])
display(studentVle_df.head(10))
```

	code_module	code_presentation	id_student	sum_click
0	AAA	2013J	28400	4
1	AAA	2013J	28400	1
2	AAA	2013J	28400	1
3	AAA	2013J	28400	11
4	AAA	2013J	28400	1
5	AAA	2013J	28400	8
6	AAA	2013J	28400	2
7	AAA	2013J	28400	15
8	AAA	2013J	28400	17
9	AAA	2013J	28400	1

In [10]:
```
studentVle_df = studentVle_df.groupby(['code_module', 'code_presentation', 'id_student
display(studentVle_df)
```

	code_module	code_presentation	id_student	sum_click
0	AAA	2013J	11391	710
1	AAA	2013J	28400	948
2	AAA	2013J	31604	1347
3	AAA	2013J	32885	796
4	AAA	2013J	38053	1303
...	...	...	...	...
12280	GGG	2014J	2620947	182
12281	GGG	2014J	2640965	41
12282	GGG	2014J	2645731	304
12283	GGG	2014J	2648187	132
12284	GGG	2014J	2684003	400

12285 rows × 4 columns

In [11]:
```
studentVle_df.shape
```

Out[11]:
```
(12285, 4)
```

Preprocess Data

In [12]:
```
display(assessment_df.head())
```

	code_module	code_presentation	id_assessment	assessment_type	date	weight
0	AAA	2013J	1752	TMA	19	10.0
1	AAA	2013J	1753	TMA	54	20.0
2	AAA	2013J	1754	TMA	117	20.0
3	AAA	2014J	1758	TMA	19	10.0
4	AAA	2014J	1759	TMA	54	20.0

In [13]:
```
#merge two tables by id_assessment.
assessment_info = pd.merge(assessment_df,studentAssessment_df,how = 'right', on = ['id

display(assessment_info.head())
```

	code_module	code_presentation	id_assessment	assessment_type	date	weight	id_student	date_
0	AAA	2013J	1752	TMA	19	10.0	11391	
1	AAA	2013J	1752	TMA	19	10.0	28400	
2	AAA	2013J	1752	TMA	19	10.0	31604	
3	AAA	2013J	1752	TMA	19	10.0	32885	
4	AAA	2013J	1752	TMA	19	10.0	38053	

In [14]:
```
student_score_df = pd.DataFrame()
student_score = {}
score_weight = {}
for i in range(assessment_info.shape[0]):
    module = assessment_info.iloc[i,0]
    presentation = assessment_info.iloc[i,1]
    student = assessment_info.iloc[i,6]
    score = assessment_info.iloc[i,9]
    weight = assessment_info.iloc[i,5]
    name = str(module) + '_' + str(presentation) + '_' + str(student)
    if name in student_score.keys():
        student_score[name] += (score * weight)
        score_weight[name] += weight
    else:
        student_score[name] = score * weight
        score_weight[name] = weight

for name in student_score.keys():
    if score_weight[name] == 0:
        student_score[name] = 0
    else:
        student_score[name] /= score_weight[name]
```

In [15]:
```
mod, pre, sid = [], [], []
for name in student_score.keys():
    temp = name.split('_')
    mod.append(temp[0])
    pre.append(temp[1])
    sid.append(temp[2])
```

In [16]:
```
student_score_df = pd.DataFrame.from_dict(student_score,orient = 'index',columns = ['a
student_score_df['code_module']= mod
student_score_df['code_presentation']= pre
student_score_df['id_student']=sid
display(student_score_df.head(10))
```

	avg_score	code_module	code_presentation	id_student
AAA_2013J_11391	81.6	AAA	2013J	11391
AAA_2013J_28400	69.2	AAA	2013J	28400
AAA_2013J_31604	72.4	AAA	2013J	31604
AAA_2013J_32885	51.0	AAA	2013J	32885
AAA_2013J_38053	73.0	AAA	2013J	38053
AAA_2013J_45462	64.8	AAA	2013J	45462
AAA_2013J_45642	72.0	AAA	2013J	45642
AAA_2013J_52130	71.6	AAA	2013J	52130
AAA_2013J_53025	77.4	AAA	2013J	53025
AAA_2013J_57506	75.2	AAA	2013J	57506

In [17]:
```
studentInfo_df['id_student'] = studentInfo_df['id_student'].astype(str)
studentVle_df['id_student'] = studentVle_df['id_student'].astype(str)
student_info = pd.merge(studentVle_df,studentInfo_df,how = 'right', on = ['code_module
```

In [18]:
```
display(student_info.head(10))
```

	code_module	code_presentation	id_student	sum_click	gender	region	highest_education	imd_
0	AAA	2013J	11391	710.0	M	East Anglian Region	HE Qualification	90-
1	AAA	2013J	28400	948.0	F	Scotland	HE Qualification	2(
2	AAA	2013J	31604	1347.0	F	South East Region	A Level or Equivalent	5(
3	AAA	2013J	32885	796.0	F	West Midlands Region	Lower Than A Level	5(
4	AAA	2013J	38053	1303.0	M	Wales	A Level or Equivalent	8(
5	AAA	2013J	45462	880.0	M	Scotland	HE Qualification	3(
6	AAA	2013J	45642	868.0	F	North Western Region	A Level or Equivalent	90-
7	AAA	2013J	52130	1054.0	F	East Anglian Region	A Level or Equivalent	7(
8	AAA	2013J	53025	1929.0	M	North Region	Post Graduate Qualification	
9	AAA	2013J	57506	973.0	M	South Region	Lower Than A Level	7(

In [19]:
```
final_df = pd.merge(student_score_df,student_info,how = 'right', on = ['code_module','
```

In [20]:
```
display(final_df.head())
```

	avg_score	code_module	code_presentation	id_student	sum_click	gender	region	highest_educ
0	81.6	AAA	2013J	11391	710.0	M	East Anglian Region	HE Qualific
1	69.2	AAA	2013J	28400	948.0	F	Scotland	HE Qualific
2	72.4	AAA	2013J	31604	1347.0	F	South East Region	A Le Equiv
3	51.0	AAA	2013J	32885	796.0	F	West Midlands Region	Lower Than A
4	73.0	AAA	2013J	38053	1303.0	M	Wales	A Le Equiv

In [21]:
```
display(final_df.describe())
```

	avg_score	sum_click	num_of_prev_attempts	studied_credits
count	11814.000000	12285.000000	12489.000000	12489.000000
mean	65.437377	970.356288	0.133157	72.926976
std	27.470409	1129.726173	0.436334	34.624923
min	0.000000	1.000000	0.000000	30.000000
25%	59.339286	261.000000	0.000000	60.000000
50%	73.569767	602.000000	0.000000	60.000000
75%	83.972141	1242.000000	0.000000	90.000000
max	100.000000	14572.000000	5.000000	430.000000

In [22]:

```
print("count of NULL values before imputation\n")
final_df.isnull().sum()
```

```
count of NULL values before imputation
```

Out[22]:

```
avg_score               675
code_module               0
code_presentation         0
id_student                0
sum_click               204
gender                    0
region                    0
highest_education         0
imd_band                422
age_band                  0
num_of_prev_attempts      0
studied_credits           0
disability                0
final_result              0
dtype: int64
```

In [23]:

```
#place the missing values of avg_score and sum_click
missing_cols = ["avg_score", "sum_click"]

for i in missing_cols:
    final_df.loc[final_df.loc[:,i].isnull(),i] = final_df.loc[:,i].median()

#since imd_band cannot be filled with mean or median, we will drop the null values.
final_df = final_df.dropna(subset=['imd_band'])

#double check the missing values of the resulting final_df
print("count of NULL values after imputation\n")
final_df.isnull().sum()
```

```
count of NULL values after imputation
```

Out[23]:
```
avg_score              0
code_module            0
code_presentation      0
id_student             0
sum_click              0
gender                 0
region                 0
highest_education      0
imd_band               0
age_band               0
num_of_prev_attempts   0
studied_credits        0
disability             0
final_result           0
dtype: int64
```

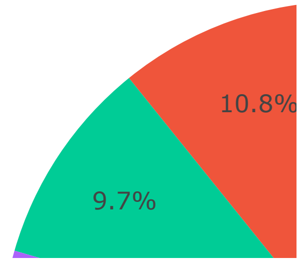
In [24]:
```python
# Create Data Labels to Visualize Data Distribution
imd = final_df["imd_band"].unique()
index = [items for items in range(len(imd))]
IMD_LABELS = dict(zip(index, imd))

gender = final_df["gender"].unique()
index1 = [items for items in range(len(gender))]
GENDER_LABELS = dict(zip(index1, gender))

edu = final_df["highest_education"].unique()
index2 = [items for items in range(len(edu))]
EDU_LABELS = dict(zip(index2, edu))

region = final_df["region"].unique()
index3 = [items for items in range(len(region))]
REGION_LABELS = dict(zip(index3, region))

age = final_df["age_band"].unique()
index4 = [items for items in range(len(age))]
AGE_LABELS = dict(zip(index3, age))


def draw_pie(df, col, title="Distribution", text_labels=None):
    '''
        Draws a Plotly pie chart from the given data.
    '''
    labels = df[col].value_counts().index.tolist()
    counts = df[col].value_counts().values.tolist()

    plot = go.Pie(labels=labels, values=counts)
    fig = go.Figure(data=[plot])
    fig.update_layout(title_text=title)
    fig.show()


draw_pie(final_df, "gender", "Gender distribution - full dataset", GENDER_LABELS)
draw_pie(final_df, "age_band", "Age band distribution - full dataset", AGE_LABELS)
draw_pie(final_df, "highest_education", "Education distribution - full dataset", EDU_L
draw_pie(final_df, "region", "Region distribution - full dataset", REGION_LABELS)
draw_pie(final_df, "imd_band", "IMD distribution - full dataset", IMD_LABELS)
```

Gender distribution - full dataset

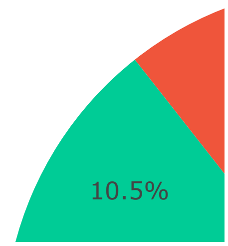
Age band distribution - full dataset

Education distribution - full dataset

Region distribution - full dataset



IMD distribution - full dataset



In [25]:
```
DROP = ["code_module", "code_presentation", "id_student"]  #these features are no long
final_df = final_df.drop(DROP, axis=1)
```

In [26]:
```
final_df.head()
```

Out[26]:

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_atte
0	81.6	710.0	M	East Anglian Region	HE Qualification	90-100%	55<=	
1	69.2	948.0	F	Scotland	HE Qualification	20-30%	35-55	
2	72.4	1347.0	F	South East Region	A Level or Equivalent	50-60%	35-55	
3	51.0	796.0	F	West Midlands Region	Lower Than A Level	50-60%	0-35	
4	73.0	1303.0	M	Wales	A Level or Equivalent	80-90%	35-55	

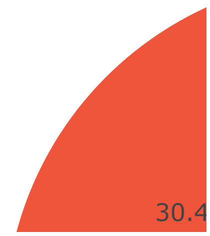
Preprocess Data

In [ ]:

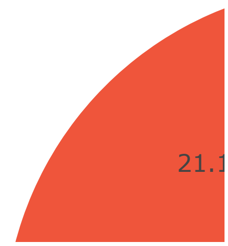
In [27]:
```
#Since proportion the group of student over 55 is so small, it will be removed from th
final_df = final_df[final_df['age_band'] != "55<="]

def replace_imd_band(x):
    '''
        This function is to reduce the number of imd_band from 10 to 5.
    '''
    if x == "0-10%" or x == "10-20":
        x = "0-20%"
    elif x == "20-30%" or x == "30-40%":
        x = "20-40%"
    elif x == "50-60%" or x == "40-50%":
        x = "40-60%"
    elif x == "60-70%" or x == "70-80%":
        x = "60-80%"
    elif x == "80-90%" or x == "90-100%":
        x = "80-100%"
    return x

final_df["imd_band"] = final_df["imd_band"].apply(lambda x: replace_imd_band(x))
```

In [28]:
```
#Show visualizations of age_band and imd_band to confirm the final_df has been changec
draw_pie(final_df, "age_band", "Age band distribution - full dataset", AGE_LABELS)
draw_pie(final_df, "imd_band", "IMD distribution - full dataset", IMD_LABELS)
```

Age band distribution - full dataset



IMD distribution - full dataset



In [29]:
```
def replace_final_results(x):
    '''
        This function converts final results from Fail, Pass, Distinct to 0(Fail) and
    '''
    if x == "Fail":
        return 0
    else:
        return 1

final_df["final_result"] = final_df["final_result"].apply(lambda x: replace_final_resu
```

In [30]:
```
def replace_prev_attempt(x):
    '''
        This function converts number of previous attempts to binary values in which N
        otherwise it returns Y.
    '''
    if x == 0:
        return "N"
    else:
        return "Y"

final_df["num_of_prev_attempts"] = final_df["num_of_prev_attempts"].apply(lambda x: re
```

In [31]:
```
display(final_df.head())
final_df.shape
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_atte
1	69.2	948.0	F	Scotland	HE Qualification	20-40%	35-55	
2	72.4	1347.0	F	South East Region	A Level or Equivalent	40-60%	35-55	
3	51.0	796.0	F	West Midlands Region	Lower Than A Level	40-60%	0-35	
4	73.0	1303.0	M	Wales	A Level or Equivalent	80-100%	35-55	
5	64.8	880.0	M	Scotland	HE Qualification	20-40%	0-35	

Out[31]:
```
(11974, 11)
```

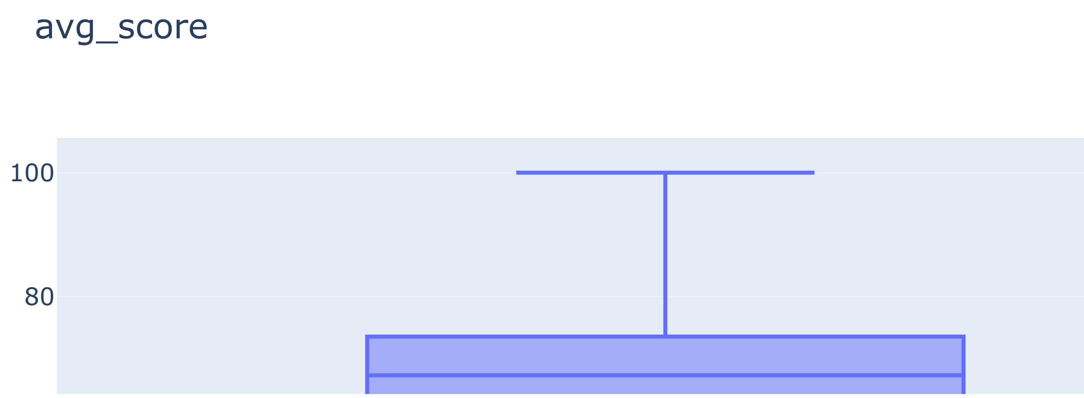
In [32]:
```
org_final_df = final_df.copy()
org_final_df = org_final_df.drop(columns=['region'])
org_final_df.head()
```

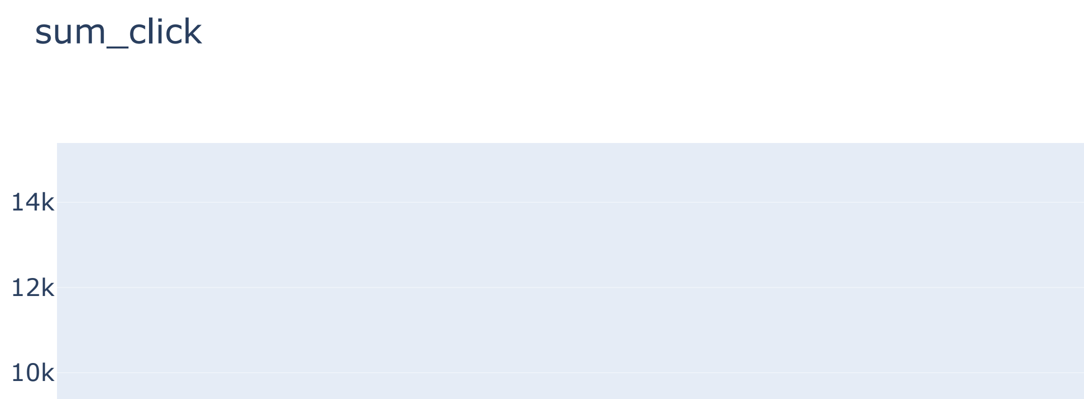
Out[32]:

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	stu
1	69.2	948.0	F	HE Qualification	20-40%	35-55	N	
2	72.4	1347.0	F	A Level or Equivalent	40-60%	35-55	N	
3	51.0	796.0	F	Lower Than A Level	40-60%	0-35	N	
4	73.0	1303.0	M	A Level or Equivalent	80-100%	35-55	N	
5	64.8	880.0	M	HE Qualification	20-40%	0-35	N	

NUMERIC FEATURES ANALYSIS

In [33]:
```
NUMERIC_FEATURES = ["avg_score", "sum_click","studied_credits"]

LABEL = "final_result"

for feature in NUMERIC_FEATURES:
    box_by_label = px.box(final_df, x=LABEL, y=feature, title=feature)
    box_by_label.show()
```


avg_score
100
80


sum_click
14k
12k
10k

In [34]:
```
# rescale function to rescale numeric features
def rescale(x, min, max):
    '''
    Rescales the value to the 0 to 1 range. Also handles outliers by replacing
    with the max value.
    '''
    return (x - min) / (max - min)

MIN_AVG_SCORE = final_df["avg_score"].min()
MAX_AVG_SCORE = final_df["avg_score"].max()
MIN_SUM_CLICK = final_df["sum_click"].min()
MAX_SUM_CLICK = final_df["sum_click"].max()
#MIN_PREV_ATMPT = final_df["num_of_prev_attempts"].min()
#MAX_PREV_ATMPT = final_df["num_of_prev_attempts"].max()
MIN_CREDIT = final_df["studied_credits"].min()
MAX_CREDIT = final_df["studied_credits"].max()

final_df["avg_score"] = final_df["avg_score"].apply(lambda x: rescale(x, MIN_AVG_SCORE
final_df["sum_click"] = final_df["sum_click"].apply(lambda x: rescale(x, MIN_SUM_CLICK
#final_df["num_of_prev_attempts"] = final_df["num_of_prev_attempts"].apply(lambda x: r
final_df["studied_credits"] = final_df["studied_credits"].apply(lambda x: rescale(x, M

display(final_df.head())
display(final_df.describe())
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_atte
1	0.692	0.064992	F	Scotland	HE Qualification	20-40%	35-55	
2	0.724	0.092375	F	South East Region	A Level or Equivalent	40-60%	35-55	
3	0.510	0.054560	F	West Midlands Region	Lower Than A Level	40-60%	0-35	
4	0.730	0.089356	M	Wales	A Level or Equivalent	80-100%	35-55	
5	0.648	0.060325	M	Scotland	HE Qualification	20-40%	0-35	

	avg_score	sum_click	studied_credits	final_result
count	11974.000000	11974.000000	11974.000000	11974.000000
mean	0.653956	0.064467	0.106942	0.708953
std	0.270206	0.075448	0.086588	0.454264
min	0.000000	0.000000	0.000000	0.000000
25%	0.600000	0.017638	0.075000	0.000000
50%	0.735698	0.040903	0.075000	1.000000
75%	0.830000	0.081583	0.150000	1.000000
max	1.000000	1.000000	1.000000	1.000000

In [35]:
```
df_pass = final_df[final_df[LABEL] == 0]
df_fail = final_df[final_df[LABEL] == 1]

SIG = 0.05
MOD_SIG = 0.1
for col in NUMERIC_FEATURES:
    pop1 = df_fail[col]
    pop2 = df_pass[col]
    stat1, p1 = normaltest(pop1)
    stat2, p2 = normaltest(pop2)
    if p1 > SIG and p2 > SIG:
        stat, p = bartlett(pop1, pop2)
        if p > SIG:
            print(col, "meets ANOVA assumptions")
        else:
            print(col, "--> Kruskal-Wallis, variance is unequal:", p)
    else:
        print(col,  "--> Kruskal-Wallis, not normally distributed:", p1, p2)
```

```
avg_score --> Kruskal-Wallis, not normally distributed: 0.0 2.5243954597350807e-158
sum_click --> Kruskal-Wallis, not normally distributed: 0.0 0.0
studied_credits --> Kruskal-Wallis, not normally distributed: 0.0 8.729033630001745e-
229
```

In [36]:
```
keep_num_feat = []
```

```
for col in NUMERIC_FEATURES:
    pop1 = df_fail[col]
    pop2 = df_pass[col]
    stat, p = f_oneway(pop1, pop2)
    if p <= SIG:
        keep_num_feat.append(col)
        print(col, "and label are not independent - keep, p =", p)
    elif p <= MOD_SIG:
        print(col, "and label may have some relationship - maybe keep, p =", p)
    else:
        print(col, "and label are independent - drop, p =", p)
```

```
avg_score and label are not independent - keep, p = 4.56970740894582e-61
sum_click and label are not independent - keep, p = 1.3070677108529551e-207
studied_credits and label are independent - drop, p = 0.2470961596211197
```

In [37]:
```
final_df = final_df.drop(columns = ["studied_credits"])
final_df.head(10)
```

Out[37]:

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_att
1	0.692	0.064992	F	Scotland	HE Qualification	20-40%	35-55	
2	0.724	0.092375	F	South East Region	A Level or Equivalent	40-60%	35-55	
3	0.510	0.054560	F	West Midlands Region	Lower Than A Level	40-60%	0-35	
4	0.730	0.089356	M	Wales	A Level or Equivalent	80-100%	35-55	
5	0.648	0.060325	M	Scotland	HE Qualification	20-40%	0-35	
6	0.720	0.059502	F	North Western Region	A Level or Equivalent	80-100%	0-35	
7	0.716	0.072267	F	East Anglian Region	A Level or Equivalent	60-80%	0-35	
9	0.752	0.066708	M	South Region	Lower Than A Level	60-80%	35-55	
10	0.714	0.075355	F	East Anglian Region	A Level or Equivalent	20-40%	0-35	
11	0.730	0.058678	M	East Anglian Region	Lower Than A Level	60-80%	35-55	

CATEGORICAL FEATURES ANALYSIS

In [38]:
```
SIG = 0.05
MOD_SIG = 0.1
```

```
CATEGORIES = ["gender", "region", "highest_education", "imd_band", "age_band", "disabi
keep = []
maybe = []

for feature in CATEGORIES:
    contingency = pd.crosstab(final_df[LABEL], final_df[feature])
    c, p, dof, expected = chi2_contingency(contingency)
    if p < SIG:
        keep.append(feature)
        print(feature, "and label are not independent - keep, p =", p)
    elif p < MOD_SIG:
        maybe.append(feature)
        print(feature, "and label may have some relationship - maybe keep, p =", p)
    else:
        print(feature, "and label are independent - drop, p =", p)
```

```
gender and label are not independent - keep, p = 0.04101756985336044
region and label are not independent - keep, p = 9.57159591725256e-19
highest_education and label are not independent - keep, p = 3.8433696941220926e-61
imd_band and label are not independent - keep, p = 3.844782925910792e-41
age_band and label are not independent - keep, p = 5.49946027840093 55e-11
disability and label are not independent - keep, p = 2.1668697218178104e-08
num_of_prev_attempts and label are not independent - keep, p = 5.805645243333483e-53
```

In [39]:
```
print("These are the kept features:")
print(keep)
```

```
These are the kept features:
['gender', 'region', 'highest_education', 'imd_band', 'age_band', 'disability', 'num_
of_prev_attempts']
```

In [40]:
```
# one hot encode the kept features
final_df_one_hot =  pd.get_dummies(final_df,
                                columns=set(keep))

#double check if the final_df has been successfully encoded
display(final_df_one_hot.head())
```

	avg_score	sum_click	final_result	gender_F	gender_M	imd_band_0-20%	imd_band_20-40%	imd_band_40-60%
1	0.692	0.064992	1	1	0	0	1	0
2	0.724	0.092375	1	1	0	0	0	1
3	0.510	0.054560	1	1	0	0	0	1
4	0.730	0.089356	1	0	1	0	0	0
5	0.648	0.060325	1	0	1	0	1	0

5 rows × 34 columns

In [41]:
```
new_features = final_df_one_hot.columns.to_list()
new_features.remove("final_result")
```

Creating Training Validation and Testing df

In [42]:
```
#Shuffle one hot encoded final_df
final_df_one_hot_shuffled = final_df_one_hot.sample(frac=1, random_state=32).reset_inc

#Create training and test df
test_size_final_df = int(len(final_df_one_hot_shuffled) * 0.2)
df_test_final = final_df_one_hot_shuffled[:test_size_final_df]
df_train_val_final = final_df_one_hot_shuffled[test_size_final_df:]

#Create training and validation df from the above training df
val_size_final = int(len(df_train_val_final) * 0.2)
df_val_final = df_train_val_final[:val_size_final]
df_train_final = df_train_val_final[val_size_final:]

df_1_x = df_train_final[new_features]
df_1_y = df_train_final[LABEL]
x_val = df_val_final[new_features]
y_val = df_val_final[LABEL]
x_test = df_test_final[new_features]
y_test = df_test_final[LABEL]
```

In [43]:
```
display(df_1_x.head())
```

	avg_score	sum_click	gender_F	gender_M	imd_band_0-20%	imd_band_20-40%	imd_band_40-60%	imd_ba
4310	0.815000	0.062178	0	1	0	0	0	
4311	0.709583	0.018187	1	0	0	1	0	
4312	0.795000	0.169789	0	1	0	0	0	
4313	0.748182	0.055315	1	0	0	0	0	
4314	0.602000	0.029579	0	1	1	0	0	

5 rows × 33 columns

Create CNN Model

In [44]:
```
def build_arbitrary_model(num_features, kernel_initializer="glorot_uniform"):
    model = keras.Sequential([
      layers.Dense(16, activation="relu", input_dim=num_features, kernel_initializer=k
      layers.Dense(16, activation="relu", kernel_initializer=kernel_initializer),
      layers.Dense(1, activation="sigmoid", kernel_initializer=kernel_initializer)
    ])
    model.compile(optimizer="adam", loss="binary_crossentropy",
                  metrics=["accuracy",
                           "TruePositives",
                           "TrueNegatives",
                           "FalsePositives",
                           "FalseNegatives"])
    return model
```

Hyperparameter tuning

In [45]:
```
def tune_model(hp):
    model = keras.Sequential()
    for i in range(hp.Int("num_layers", min_value=1, max_value=4, step=1)):
        model.add(layers.Dense(units=hp.Int("units_" + str(i), min_value=57, max_value

    model.add(layers.Dense(1, activation="sigmoid"))
    model.compile(optimizer="adam", loss="binary_crossentropy",
                  metrics=["accuracy", "TruePositives", "TrueNegatives",
                           "FalsePositives", "FalseNegatives"])
    return model

tuner_acc = RandomSearch(
    tune_model, objective="val_accuracy",
    max_trials=10, executions_per_trial=3, project_name="project_3"
)

tuner_acc.search_space_summary()

tuner_acc.search(df_1_x.values,
                 df_1_y.values,
                 epochs=100,
                 batch_size=600,
                 validation_data=(x_val.values, y_val.values),
                 callbacks=[keras.callbacks.EarlyStopping(monitor="val_loss", patience

tuner_acc.results_summary()
```

Best Optimized Model

In [46]:
```
def best_model(layer_nums):
    model = keras.Sequential()
    for i in range(len(layer_nums)):
        model.add(layers.Dense(layer_nums[i], activation="relu"))

    model.add(layers.Dense(1, activation="sigmoid"))
    model.compile(optimizer="adam", loss="binary_crossentropy",
                  metrics=["accuracy",
                           "TruePositives", "TrueNegatives",
                           "FalsePositives", "FalseNegatives"])
    return model
```

In [47]:
```
best_model_acc = best_model([95, 76, 95, 76])
best_model_acc.fit(df_1_x.values, df_1_y.values, epochs=100, batch_size=600,
                   validation_data=(x_val.values, y_val.values),
                   callbacks=[keras.callbacks.EarlyStopping(monitor="val_loss", patien
```

```
Epoch 1/100
13/13 [==============================] - ETA: 0s - loss: 0.6789 - accuracy: 0.6383 -
true_positives: 344.0000 - true_negatives: 39.0000 - false_positives: 136.0000 - fals
e_negatives: 81.00 - 0s 35ms/step - loss: 0.6085 - accuracy: 0.7067 - true_positives:
5377.0000 - true_negatives: 39.0000 - false_positives: 2167.0000 - false_negatives: 8
1.0000 - val_loss: 0.5855 - val_accuracy: 0.7098 - val_true_positives: 1360.0000 - va
l_true_negatives: 0.0000e+00 - val_false_positives: 556.0000 - val_false_negatives:
0.0000e+00
Epoch 2/100
13/13 [==============================] - ETA: 0s - loss: 0.5616 - accuracy: 0.7350 -
true_positives: 441.0000 - true_negatives: 0.0000e+00 - false_positives: 159.0000 - f
alse_negatives: 0.0000e+ - 0s 3ms/step - loss: 0.5766 - accuracy: 0.7122 - true_posit
ives: 5458.0000 - true_negatives: 0.0000e+00 - false_positives: 2206.0000 - false_neg
atives: 0.0000e+00 - val_loss: 0.5710 - val_accuracy: 0.7098 - val_true_positives: 13
60.0000 - val_true_negatives: 0.0000e+00 - val_false_positives: 556.0000 - val_false_
negatives: 0.0000e+00
Epoch 3/100
13/13 [==============================] - ETA: 0s - loss: 0.5831 - accuracy: 0.7067 -
true_positives: 424.0000 - true_negatives: 0.0000e+00 - false_positives: 176.0000 - f
alse_negatives: 0.0000e+ - 0s 3ms/step - loss: 0.5625 - accuracy: 0.7172 - true_posit
ives: 5420.0000 - true_negatives: 77.0000 - false_positives: 2129.0000 - false_negati
ves: 38.0000 - val_loss: 0.5644 - val_accuracy: 0.7150 - val_true_positives: 1321.000
0 - val_true_negatives: 49.0000 - val_false_positives: 507.0000 - val_false_negative
s: 39.0000
Epoch 4/100
13/13 [==============================] - ETA: 0s - loss: 0.5378 - accuracy: 0.7567 -
true_positives: 433.0000 - true_negatives: 21.0000 - false_positives: 132.0000 - fals
e_negatives: 14.00 - 0s 3ms/step - loss: 0.5535 - accuracy: 0.7234 - true_positives:
5266.0000 - true_negatives: 278.0000 - false_positives: 1928.0000 - false_negatives:
192.0000 - val_loss: 0.5623 - val_accuracy: 0.7129 - val_true_positives: 1288.0000 -
val_true_negatives: 78.0000 - val_false_positives: 478.0000 - val_false_negatives: 7
2.0000
Epoch 5/100
13/13 [==============================] - ETA: 0s - loss: 0.5643 - accuracy: 0.7100 -
true_positives: 392.0000 - true_negatives: 34.0000 - false_positives: 150.0000 - fals
e_negatives: 24.00 - 0s 2ms/step - loss: 0.5446 - accuracy: 0.7251 - true_positives:
5130.0000 - true_negatives: 427.0000 - false_positives: 1779.0000 - false_negatives:
328.0000 - val_loss: 0.5596 - val_accuracy: 0.7156 - val_true_positives: 1273.0000 -
val_true_negatives: 98.0000 - val_false_positives: 458.0000 - val_false_negatives: 8
7.0000
Epoch 6/100
13/13 [==============================] - ETA: 0s - loss: 0.5303 - accuracy: 0.7300 -
true_positives: 400.0000 - true_negatives: 38.0000 - false_positives: 139.0000 - fals
e_negatives: 23.00 - 0s 3ms/step - loss: 0.5373 - accuracy: 0.7324 - true_positives:
5129.0000 - true_negatives: 484.0000 - false_positives: 1722.0000 - false_negatives:
329.0000 - val_loss: 0.5528 - val_accuracy: 0.7176 - val_true_positives: 1266.0000 -
val_true_negatives: 109.0000 - val_false_positives: 447.0000 - val_false_negatives: 9
4.0000
Epoch 7/100
13/13 [==============================] - ETA: 0s - loss: 0.5204 - accuracy: 0.7550 -
true_positives: 408.0000 - true_negatives: 45.0000 - false_positives: 125.0000 - fals
e_negatives: 22.00 - 0s 3ms/step - loss: 0.5294 - accuracy: 0.7343 - true_positives:
5115.0000 - true_negatives: 513.0000 - false_positives: 1693.0000 - false_negatives:
343.0000 - val_loss: 0.5470 - val_accuracy: 0.7208 - val_true_positives: 1258.0000 -
val_true_negatives: 123.0000 - val_false_positives: 433.0000 - val_false_negatives: 1
02.0000
Epoch 8/100
13/13 [==============================] - ETA: 0s - loss: 0.4926 - accuracy: 0.7850 -
true_positives: 425.0000 - true_negatives: 46.0000 - false_positives: 113.0000 - fals
e_negatives: 16.00 - 0s 2ms/step - loss: 0.5203 - accuracy: 0.7381 - true_positives:
```

```
5057.0000 - true_negatives: 600.0000 - false_positives: 1606.0000 - false_negatives:
401.0000 - val_loss: 0.5372 - val_accuracy: 0.7171 - val_true_positives: 1225.0000 -
val_true_negatives: 149.0000 - val_false_positives: 407.0000 - val_false_negatives: 1
35.0000
Epoch 9/100
13/13 [==============================] - ETA: 0s - loss: 0.5147 - accuracy: 0.7433 -
true_positives: 385.0000 - true_negatives: 61.0000 - false_positives: 121.0000 - fals
e_negatives: 33.00 - 0s 2ms/step - loss: 0.5089 - accuracy: 0.7413 - true_positives:
5086.0000 - true_negatives: 595.0000 - false_positives: 1611.0000 - false_negatives:
372.0000 - val_loss: 0.5262 - val_accuracy: 0.7223 - val_true_positives: 1239.0000 -
val_true_negatives: 145.0000 - val_false_positives: 411.0000 - val_false_negatives: 1
21.0000
Epoch 10/100
13/13 [==============================] - ETA: 0s - loss: 0.5347 - accuracy: 0.7317 -
true_positives: 381.0000 - true_negatives: 58.0000 - false_positives: 130.0000 - fals
e_negatives: 31.00 - 0s 3ms/step - loss: 0.5002 - accuracy: 0.7452 - true_positives:
4984.0000 - true_negatives: 727.0000 - false_positives: 1479.0000 - false_negatives:
474.0000 - val_loss: 0.5361 - val_accuracy: 0.7255 - val_true_positives: 1288.0000 -
val_true_negatives: 102.0000 - val_false_positives: 454.0000 - val_false_negatives: 7
2.0000
Epoch 11/100
13/13 [==============================] - ETA: 0s - loss: 0.5122 - accuracy: 0.7467 -
true_positives: 402.0000 - true_negatives: 46.0000 - false_positives: 133.0000 - fals
e_negatives: 19.00 - 0s 3ms/step - loss: 0.4913 - accuracy: 0.7548 - true_positives:
5026.0000 - true_negatives: 759.0000 - false_positives: 1447.0000 - false_negatives:
432.0000 - val_loss: 0.5112 - val_accuracy: 0.7427 - val_true_positives: 1225.0000 -
val_true_negatives: 198.0000 - val_false_positives: 358.0000 - val_false_negatives: 1
35.0000
Epoch 12/100
13/13 [==============================] - ETA: 0s - loss: 0.5336 - accuracy: 0.7267 -
true_positives: 384.0000 - true_negatives: 52.0000 - false_positives: 106.0000 - fals
e_negatives: 58.00 - 0s 3ms/step - loss: 0.4801 - accuracy: 0.7590 - true_positives:
4955.0000 - true_negatives: 862.0000 - false_positives: 1344.0000 - false_negatives:
503.0000 - val_loss: 0.5093 - val_accuracy: 0.7396 - val_true_positives: 1245.0000 -
val_true_negatives: 172.0000 - val_false_positives: 384.0000 - val_false_negatives: 1
15.0000
Epoch 13/100
13/13 [==============================] - ETA: 0s - loss: 0.4610 - accuracy: 0.7650 -
true_positives: 389.0000 - true_negatives: 70.0000 - false_positives: 104.0000 - fals
e_negatives: 37.00 - 0s 3ms/step - loss: 0.4732 - accuracy: 0.7663 - true_positives:
4943.0000 - true_negatives: 930.0000 - false_positives: 1276.0000 - false_negatives:
515.0000 - val_loss: 0.5192 - val_accuracy: 0.7416 - val_true_positives: 1268.0000 -
val_true_negatives: 153.0000 - val_false_positives: 403.0000 - val_false_negatives: 9
2.0000
Epoch 14/100
13/13 [==============================] - ETA: 0s - loss: 0.5066 - accuracy: 0.7617 -
true_positives: 398.0000 - true_negatives: 59.0000 - false_positives: 118.0000 - fals
e_negatives: 25.00 - 0s 3ms/step - loss: 0.4743 - accuracy: 0.7687 - true_positives:
4922.0000 - true_negatives: 969.0000 - false_positives: 1237.0000 - false_negatives:
536.0000 - val_loss: 0.5172 - val_accuracy: 0.7422 - val_true_positives: 1270.0000 -
val_true_negatives: 152.0000 - val_false_positives: 404.0000 - val_false_negatives: 9
0.0000
Epoch 15/100
13/13 [==============================] - ETA: 0s - loss: 0.4656 - accuracy: 0.7683 -
true_positives: 401.0000 - true_negatives: 60.0000 - false_positives: 112.0000 - fals
e_negatives: 27.00 - 0s 3ms/step - loss: 0.4667 - accuracy: 0.7745 - true_positives:
4920.0000 - true_negatives: 1016.0000 - false_positives: 1190.0000 - false_negatives:
538.0000 - val_loss: 0.5039 - val_accuracy: 0.7516 - val_true_positives: 1199.0000 -
val_true_negatives: 241.0000 - val_false_positives: 315.0000 - val_false_negatives: 1
61.0000
```

```
Epoch 16/100
13/13 [==============================] - ETA: 0s - loss: 0.4731 - accuracy: 0.7550 -
true_positives: 364.0000 - true_negatives: 89.0000 - false_positives: 86.0000 - false
_negatives: 61.000 - 0s 3ms/step - loss: 0.4626 - accuracy: 0.7744 - true_positives:
4960.0000 - true_negatives: 975.0000 - false_positives: 1231.0000 - false_negatives:
498.0000 - val_loss: 0.5003 - val_accuracy: 0.7547 - val_true_positives: 1222.0000 -
val_true_negatives: 224.0000 - val_false_positives: 332.0000 - val_false_negatives: 1
38.0000
Epoch 17/100
13/13 [==============================] - ETA: 0s - loss: 0.4602 - accuracy: 0.7717 -
true_positives: 384.0000 - true_negatives: 79.0000 - false_positives: 90.0000 - false
_negatives: 47.000 - 0s 3ms/step - loss: 0.4566 - accuracy: 0.7781 - true_positives:
4890.0000 - true_negatives: 1073.0000 - false_positives: 1133.0000 - false_negatives:
568.0000 - val_loss: 0.5031 - val_accuracy: 0.7500 - val_true_positives: 1235.0000 -
val_true_negatives: 202.0000 - val_false_positives: 354.0000 - val_false_negatives: 1
25.0000
Epoch 18/100
13/13 [==============================] - ETA: 0s - loss: 0.4631 - accuracy: 0.7850 -
true_positives: 409.0000 - true_negatives: 62.0000 - false_positives: 97.0000 - false
_negatives: 32.000 - 0s 3ms/step - loss: 0.4509 - accuracy: 0.7850 - true_positives:
4956.0000 - true_negatives: 1060.0000 - false_positives: 1146.0000 - false_negatives:
502.0000 - val_loss: 0.5028 - val_accuracy: 0.7610 - val_true_positives: 1179.0000 -
val_true_negatives: 279.0000 - val_false_positives: 277.0000 - val_false_negatives: 1
81.0000
Epoch 19/100
13/13 [==============================] - ETA: 0s - loss: 0.4349 - accuracy: 0.7967 -
true_positives: 397.0000 - true_negatives: 81.0000 - false_positives: 62.0000 - false
_negatives: 60.000 - 0s 3ms/step - loss: 0.4533 - accuracy: 0.7829 - true_positives:
4926.0000 - true_negatives: 1074.0000 - false_positives: 1132.0000 - false_negatives:
532.0000 - val_loss: 0.5032 - val_accuracy: 0.7557 - val_true_positives: 1169.0000 -
val_true_negatives: 279.0000 - val_false_positives: 277.0000 - val_false_negatives: 1
91.0000
```

Out[47]:
```
<tensorflow.python.keras.callbacks.History at 0x1a1cb58d508>
```

Precisions, Recalls and PPVs

In [48]:
```
print("Overall accuracy:")
best_model_acc.evaluate(x_test.values, y_test.values)[1]
```

```
Overall accuracy:
75/75 [==============================] - ETA: 0s - loss: 0.4030 - accuracy: 0.8125 -
true_positives: 19.0000 - true_negatives: 7.0000 - false_positives: 2.0000 - false_ne
gatives: 4.000 - 0s 600us/step - loss: 0.5006 - accuracy: 0.7419 - true_positives: 14
28.0000 - true_negatives: 348.0000 - false_positives: 375.0000 - false_negatives: 24
3.0000
```

Out[48]:
```
0.7418546080589294
```

In [49]:
```
def precision_recall(tp, tn, fp, fn):
    precision = tp / (tp + fp)
    recall = tp / (tp + fn)
    return precision, recall
```

In [50]:
```
y_test_b = df_test_final[df_test_final["gender_F"] == 1][LABEL]
print()
pred_f = best_model_acc.evaluate(x_test[x_test["gender_F"] == 1].values, y_test_b.valu
precision_f, recall_f = precision_recall(pred_f[2], pred_f[3], pred_f[4], pred_f[5])
```

```
print("Accuracy of Female:".format(pred_f[1]))
print("Precision of Male: {}, recall of Male: {}".format(precision_f, recall_f))
```

```
 1/37 [..............................] - ETA: 0s - loss: 0.5309 - accuracy: 0.7500 -
true_positives: 19.0000 - true_negatives: 5.0000 - false_positives: 5.0000 - false_ne
gatives: 3.0000WARNING:tensorflow:Callbacks method `on_test_batch_end` is slow compar
ed to the batch time (batch time: 0.0000s vs `on_test_batch_end` time: 0.0010s). Chec
k your callbacks.
37/37 [==============================] - 0s 622us/step - loss: 0.5239 - accuracy: 0.7
286 - true_positives: 714.0000 - true_negatives: 145.0000 - false_positives: 203.0000
- false_negatives: 117.0000
Accuracy of Female:
Precision of Male: 0.7786259541984732, recall of Male: 0.8592057761732852
```

In [51]:

```
y_test_b = df_test_final[df_test_final["gender_M"] == 1][LABEL]

pred_m = best_model_acc.evaluate(x_test[x_test["gender_M"] == 1].values, y_test_b.valu
precision_m, recall_m = precision_recall(pred_m[2], pred_m[3], pred_m[4], pred_m[5])
print("Accuracy of Male:".format(pred_m[1]))
print("Precision of Male: {}, recall of Male: {}".format(precision_m, recall_m))
```

```
38/38 [==============================] - ETA: 0s - loss: 0.3837 - accuracy: 0.8125 -
true_positives: 19.0000 - true_negatives: 7.0000 - false_positives: 3.0000 - false_ne
gatives: 3.000 - 0s 737us/step - loss: 0.4781 - accuracy: 0.7547 - true_positives: 71
4.0000 - true_negatives: 203.0000 - false_positives: 172.0000 - false_negatives: 126.
0000
Accuracy of Male:
Precision of Male: 0.8058690744920993, recall of Male: 0.85
```

In [52]:

```
y_test_b = df_test_final[df_test_final["age_band_0-35"] == 1][LABEL]

pred_0_35 = best_model_acc.evaluate(x_test[x_test["age_band_0-35"] == 1].values, y_tes
precision_0_35, recall_0_35 = precision_recall(pred_0_35[2], pred_0_35[3], pred_0_35[4
print("Accuracy of Age Band between 0 and 35:".format(pred_0_35[1]))
print("Precision and recall of age between 0 and 35: {}, {}".format(precision_0_35, re
```

```
 1/52 [..............................] - ETA: 0s - loss: 0.4691 - accuracy: 0.7812 -
true_positives: 18.0000 - true_negatives: 7.0000 - false_positives: 2.0000 - false_ne
gatives: 5.0000WARNING:tensorflow:Callbacks method `on_test_batch_begin` is slow comp
ared to the batch time (batch time: 0.0000s vs `on_test_batch_begin` time: 0.0010s).
Check your callbacks.
52/52 [==============================] - 0s 539us/step - loss: 0.5037 - accuracy: 0.7
376 - true_positives: 944.0000 - true_negatives: 273.0000 - false_positives: 252.0000
- false_negatives: 181.0000
Accuracy of Age Band between 0 and 35:
Precision and recall of age between 0 and 35: 0.7892976588628763, 0.8391111111111111
```

In [53]:

```
y_test_3555 = df_test_final[df_test_final["age_band_35-55"] == 1][LABEL]

pred_35_55 = best_model_acc.evaluate(x_test[x_test["age_band_35-55"] == 1].values, y_t
precision_35_55, recall_35_55 = precision_recall(pred_35_55[2], pred_35_55[3], pred_35
print("Accuracy of Age Band between 35 and 55: {}".format(pred_35_55[1]))
print("Precision and recall of Age band between 35 and 55: {} {}".format(precision_35_
```

```
24/24 [==============================] - ETA: 0s - loss: 0.6023 - accuracy: 0.7188 -
true_positives: 18.0000 - true_negatives: 5.0000 - false_positives: 7.0000 - false_ne
gatives: 2.000 - 0s 583us/step - loss: 0.4938 - accuracy: 0.7513 - true_positives: 48
4.0000 - true_negatives: 75.0000 - false_positives: 123.0000 - false_negatives: 62.00
00
Accuracy of Age Band between 35 and 55: 0.7513440847396851
Precision and recall of Age band between 35 and 55: 0.7973640856672158 0.886446886446
8864
```

In [54]:

```
y_test_b = df_test_final[df_test_final["imd_band_0-20%"] == 1][LABEL]

pred_0_20 = best_model_acc.evaluate(x_test[x_test["imd_band_0-20%"] == 1].values, y_te
precision_0_20, recall_0_20 = precision_recall(pred_0_20[2], pred_0_20[3], pred_0_20[4
print("Accuracy of IMD Band between 0 and 20: {}".format(pred_0_20[1]))
print("Precision and recall of IMD band between 0 and 20: {} {}".format(precision_0_20
```

```
15/15 [==============================] - ETA: 0s - loss: 0.7924 - accuracy: 0.4688 -
true_positives: 8.0000 - true_negatives: 7.0000 - false_positives: 9.0000 - false_neg
atives: 8.00 - 0s 711us/step - loss: 0.5537 - accuracy: 0.7143 - true_positives: 212.
0000 - true_negatives: 128.0000 - false_positives: 72.0000 - false_negatives: 64.0000
Accuracy of IMD Band between 0 and 20: 0.7142857313156128
Precision and recall of IMD band between 0 and 20: 0.7464788732394366 0.7681159420289
855
```

In [55]:

```
y_test_b = df_test_final[df_test_final["imd_band_20-40%"] == 1][LABEL]
print()
pred_20_40 = best_model_acc.evaluate(x_test[x_test["imd_band_20-40%"] == 1].values, y_
precision_20_40, recall_20_40 = precision_recall(pred_20_40[2], pred_20_40[3], pred_20
print("Accuracy of IMD Band between 20 and 40: {}".format(pred_20_40[1]))
print("Precision and recall of IMD band between 20 and 40: {}, {}".format(precision_20
```

```
17/17 [==============================] - ETA: 0s - loss: 0.4335 - accuracy: 0.8438 -
true_positives: 17.0000 - true_negatives: 10.0000 - false_positives: 2.0000 - false_n
egatives: 3.00 - 0s 647us/step - loss: 0.5092 - accuracy: 0.7357 - true_positives: 30
4.0000 - true_negatives: 83.0000 - false_positives: 82.0000 - false_negatives: 57.000
0
Accuracy of IMD Band between 20 and 40: 0.7357414364814758
Precision and recall of IMD band between 20 and 40: 0.7875647668393783, 0.84210526315
78947
```

In [56]:

```
y_test_b = df_test_final[df_test_final["imd_band_40-60%"] == 1][LABEL]

pred_40_60 = best_model_acc.evaluate(x_test[x_test["imd_band_40-60%"] == 1].values, y_
precision_40_60, recall_40_60 = precision_recall(pred_40_60[2], pred_40_60[3], pred_40
print("Accuracy of IMD Band between 40 and 60: {}".format(pred_40_60[1]))
print("Precision and recall of IMD band between 40 and 60: {}, {}".format(precision_40
```

```
17/17 [==============================] - ETA: 0s - loss: 0.4934 - accuracy: 0.6562 -
true_positives: 18.0000 - true_negatives: 3.0000 - false_positives: 6.0000 - false_ne
gatives: 5.000 - 0s 656us/step - loss: 0.5017 - accuracy: 0.7273 - true_positives: 31
3.0000 - true_negatives: 63.0000 - false_positives: 84.0000 - false_negatives: 57.000
0
Accuracy of IMD Band between 40 and 60: 0.7272727489471436
Precision and recall of IMD band between 40 and 60: 0.7884130982367759, 0.84594594594
5946
```

In [57]:

```
y_test_b = df_test_final[df_test_final["imd_band_60-80%"] == 1][LABEL]

pred_60_80 = best_model_acc.evaluate(x_test[x_test["imd_band_60-80%"] == 1].values, y_
precision_60_80, recall_60_80 = precision_recall(pred_60_80[2], pred_60_80[3], pred_60
```

```
print("Accuracy of IMD Band between 60 and 80: {}".format(pred_60_80[1]))
print("Precision and recall of IMD band between 60 and 80: {}, {}".format(precision_6
```

```
15/15 [==============================] - ETA: 0s - loss: 0.3974 - accuracy: 0.7812 -
true_positives: 20.0000 - true_negatives: 5.0000 - false_positives: 4.0000 - false_ne
gatives: 3.000 - 0s 733us/step - loss: 0.4950 - accuracy: 0.7505 - true_positives: 30
2.0000 - true_negatives: 56.0000 - false_positives: 77.0000 - false_negatives: 42.000
0
Accuracy of IMD Band between 60 and 80: 0.75052410364151
Precision and recall of IMD band between 60 and 80: 0.7968337730870713, 0.87790697674
4186
```

In [58]:

```
y_test_b = df_test_final[df_test_final["imd_band_80-100%"] == 1][LABEL]

pred_80_100 = best_model_acc.evaluate(x_test[x_test["imd_band_80-100%"] == 1].values,
precision_80_100, recall_80_100 = precision_recall(pred_80_100[2], pred_80_100[3], pre
print("Accuracy of IMD Band between 60 and 80: {}".format(pred_80_100[1]))
print("Precision and recall of IMD band between 60 and 80: {}, {}".format(precision_8
```

```
13/13 [==============================] - ETA: 0s - loss: 0.4236 - accuracy: 0.8438 -
true_positives: 26.0000 - true_negatives: 1.0000 - false_positives: 4.0000 - false_ne
gatives: 1.000 - 0s 769us/step - loss: 0.4313 - accuracy: 0.7915 - true_positives: 29
7.0000 - true_negatives: 18.0000 - false_positives: 60.0000 - false_negatives: 23.000
0
Accuracy of IMD Band between 60 and 80: 0.7914572954177856
Precision and recall of IMD band between 60 and 80: 0.8319327731092437, 0.928125
```

In [59]:

```
imd_band_precisions = [precision_0_20,precision_20_40, precision_40_60, precision_60_8
age_band_precisions = [precision_0_35,precision_35_55]
gender_precisions = [precision_m,precision_f]

def ppv_diff(imd_list):
    result = 0
    ppv_list = []
    for i in range(len(imd_list)):
        mean_1 = imd_list[i]
        for j in range(i + 1, len(imd_list)):
            mean_2 = imd_list[j]
            ppv_list.append(abs(mean_1 - mean_2))
    result = sum(ppv_list)/len(ppv_list)
    return result

ppv_imd_band = ppv_diff(imd_band_precisions)
ppv_age_band = ppv_diff(age_band_precisions)
ppv_gender = ppv_diff(gender_precisions)

print("ppv_imd_band: {}".format(ppv_imd_band))
print("ppv_age_band: {}".format(ppv_age_band))
print("ppv_gender: {}".format(ppv_gender))
```

```
ppv_imd_band: 0.036035361197461445
ppv_age_band: 0.008066426804339555
ppv_gender: 0.027243120293626077
```

The above average ppv results are below 0.05 indicating that my model is fair when it is applied to predict gender, age_band and imd_band.

LIME

In [60]:
```
def preprocess_df(df, numeric_features, binary_features, scale, categories):

    new_df = df.copy(deep=True)
    for feature in df.columns:
        if feature in numeric_features:
          # For some reason, there are numeric features encoded as strings in the test
            new_df[feature] = pd.to_numeric(df[feature])
            new_df[feature] = new_df[feature].apply(lambda x: rescale(x, scale[feature
        elif feature in binary_features:
            new_df[feature] = pd.get_dummies(df[feature], drop_first=True) # Binary en
        else: # manually encode categorical features
            for value in categories[feature]:
                new_df[feature + "_" + value] = df[feature].apply(lambda x: 1 if x ==
            new_df = new_df.drop(feature, axis=1) # drop the original column
    return new_df


def split(pre_split_df):

    train_size = int(len(pre_split_df) * 0.8)
    return pre_split_df[:train_size], pre_split_df[train_size:]



def preprocess_df_for_lime(df, numeric_features, binary_features, scale, categories):

    new_df = df.copy(deep=True)

    for feature in df.columns:
        if feature in numeric_features:
            # For some reason, there are numeric features encoded as strings in the te
            new_df[feature] = pd.to_numeric(df[feature])
            new_df[feature] = new_df[feature].apply(lambda x: rescale(x, scale[feature
        elif feature in binary_features:
            new_df[feature] = pd.get_dummies(df[feature], drop_first=True) # Binary en
        else: # manually encode categorical as levels instead of dummy encoding, which
            new_df[feature] = df[feature].apply(lambda x: categories[feature].index(x)
    return new_df

def convert_lime_df_to_keras(df, categories, binary):
    new_df = df.copy(deep=True)
    for feature in df.columns:
        if feature in categories and feature not in binary:
            for i, value in enumerate(categories[feature]):
                new_df[feature + "_" + value] = df[feature].apply(lambda x: 1 if x ==
            new_df = new_df.drop(feature, axis=1)
    return new_df

def build_arbitrary_model1():
    model = keras.Sequential([
      layers.Dense(16, activation="relu"),
      layers.Dense(16, activation="relu"),
      layers.Dense(1, activation="sigmoid")
  ])
    model.compile(optimizer="adam", loss="binary_crossentropy",
                  metrics=["accuracy"])
    return model
```

In [61]:
```
FAIL = 0
```

```
PASS = 1

# These are the features kept for this notebook
features = final_df.columns.tolist()

numeric_features = ["avg_score", "sum_click"]

categorical_features = ["region", "highest_education", "imd_band", "age_band", "num_of

binary = ["gender", "disability", LABEL]
```

In [62]:
```
final_df_shuffled = final_df.sample(frac=1, random_state=32).reset_index(drop=True)
test_size_final_df1 = int(len(final_df_shuffled) * 0.2)
df_test_final1 = final_df_shuffled[:test_size_final_df1]
df_train_val_final1 = final_df_shuffled[test_size_final_df1:]
val_size_final1 = int(len(df_train_val_final1) * 0.2)
df_val_final1 = df_train_val_final1[:val_size_final1]
df_train_final1 = df_train_val_final1[val_size_final1:]
```

In [63]:
```
num_values = {}

for feature in numeric_features:
    num_values[feature] = {
        "min": df_train_final1[feature].min(),
        "max": df_train_final1[feature].max()
    }

for k,v in num_values.items():
    print(k, v)
```

```
avg_score {'min': 0.0, 'max': 1.0}
sum_click {'min': 0.0, 'max': 1.0}
```

In [64]:
```
cat_values = {}
for feature in categorical_features:
    cat_values[feature] = list(df_train_final1[feature].value_counts().index)
cat_values["gender"] = ["F", "M"]
cat_values["disability"] = ["Y", "N"]

for k,v in cat_values.items():
    print(k, v)
```

```
region ['Scotland', 'East Anglian Region', 'South Region', 'London Region', 'North We
stern Region', 'South West Region', 'West Midlands Region', 'Wales', 'East Midlands R
egion', 'South East Region', 'Yorkshire Region', 'Ireland', 'North Region']
highest_education ['A Level or Equivalent', 'Lower Than A Level', 'HE Qualification',
'No Formal quals', 'Post Graduate Qualification']
imd_band ['20-40%', '40-60%', '0-20%', '60-80%', '80-100%']
age_band ['0-35', '35-55']
num_of_prev_attempts ['N', 'Y']
gender ['F', 'M']
disability ['Y', 'N']
```

In [65]:
```
df_train_final_lime = preprocess_df(df_train_final1[features], numeric_features, binar
                              num_values, cat_values)
df_val_final_lime = preprocess_df(df_val_final1[features], numeric_features, binary,
                            num_values, cat_values)

x_train = df_train_final_lime.drop(LABEL, axis=1)
y_train = df_train_final_lime[LABEL]
```

```
x_val = df_val_final_lime.drop(LABEL, axis=1)
y_val = df_val_final_lime[LABEL]
```

In [66]:
```
print((x_train.shape))
display(x_train.head())
```

```
(7664, 31)
```

	avg_score	sum_click	gender	disability	region_Scotland	region_East Anglian Region	region_South Region	region_Lo Re
4310	0.815000	0.062178	1	0	0	0	0	
4311	0.709583	0.018187	0	0	0	0	0	
4312	0.795000	0.169789	1	0	0	0	0	
4313	0.748182	0.055315	0	0	0	0	0	
4314	0.602000	0.029579	1	0	0	0	0	

5 rows × 31 columns

In [67]:
```
model1 = best_model([95,76, 95, 76])
model1.fit(x_train.values, y_train.values,
          validation_data=(x_val.values, y_val.values),
          epochs=100, batch_size=600,
          callbacks=[keras.callbacks.EarlyStopping(monitor="loss", patience=3)])
```

```
Epoch 1/100
13/13 [==============================] - ETA: 0s - loss: 0.6773 - accuracy: 0.6967 -
true_positives: 413.0000 - true_negatives: 5.0000 - false_positives: 170.0000 - false
_negatives: 12.000 - 1s 43ms/step - loss: 0.6096 - accuracy: 0.7112 - true_positives:
5446.0000 - true_negatives: 5.0000 - false_positives: 2201.0000 - false_negatives: 1
2.0000 - val_loss: 0.5943 - val_accuracy: 0.7098 - val_true_positives: 1360.0000 - va
l_true_negatives: 0.0000e+00 - val_false_positives: 556.0000 - val_false_negatives:
0.0000e+00
Epoch 2/100
13/13 [==============================] - ETA: 0s - loss: 0.5634 - accuracy: 0.7350 -
true_positives: 441.0000 - true_negatives: 0.0000e+00 - false_positives: 159.0000 - f
alse_negatives: 0.0000e+ - 0s 3ms/step - loss: 0.5804 - accuracy: 0.7122 - true_posit
ives: 5458.0000 - true_negatives: 0.0000e+00 - false_positives: 2206.0000 - false_neg
atives: 0.0000e+00 - val_loss: 0.5777 - val_accuracy: 0.7098 - val_true_positives: 13
60.0000 - val_true_negatives: 0.0000e+00 - val_false_positives: 556.0000 - val_false_
negatives: 0.0000e+00
Epoch 3/100
13/13 [==============================] - ETA: 0s - loss: 0.5840 - accuracy: 0.7067 -
true_positives: 424.0000 - true_negatives: 0.0000e+00 - false_positives: 176.0000 - f
alse_negatives: 0.0000e+ - 0s 3ms/step - loss: 0.5682 - accuracy: 0.7123 - true_posit
ives: 5458.0000 - true_negatives: 1.0000 - false_positives: 2205.0000 - false_negativ
es: 0.0000e+00 - val_loss: 0.5696 - val_accuracy: 0.7135 - val_true_positives: 1359.0
000 - val_true_negatives: 8.0000 - val_false_positives: 548.0000 - val_false_negative
s: 1.0000
Epoch 4/100
13/13 [==============================] - ETA: 0s - loss: 0.5395 - accuracy: 0.7483 -
true_positives: 446.0000 - true_negatives: 3.0000 - false_positives: 150.0000 - false
_negatives: 1.00 - 0s 3ms/step - loss: 0.5587 - accuracy: 0.7195 - true_positives: 53
94.0000 - true_negatives: 120.0000 - false_positives: 2086.0000 - false_negatives: 6
4.0000 - val_loss: 0.5648 - val_accuracy: 0.7119 - val_true_positives: 1309.0000 - va
l_true_negatives: 55.0000 - val_false_positives: 501.0000 - val_false_negatives: 51.0
000
Epoch 5/100
13/13 [==============================] - ETA: 0s - loss: 0.5732 - accuracy: 0.6917 -
true_positives: 397.0000 - true_negatives: 18.0000 - false_positives: 166.0000 - fals
e_negatives: 19.00 - 0s 3ms/step - loss: 0.5501 - accuracy: 0.7243 - true_positives:
5204.0000 - true_negatives: 347.0000 - false_positives: 1859.0000 - false_negatives:
254.0000 - val_loss: 0.5599 - val_accuracy: 0.7156 - val_true_positives: 1290.0000 -
val_true_negatives: 81.0000 - val_false_positives: 475.0000 - val_false_negatives: 7
0.0000
Epoch 6/100
13/13 [==============================] - ETA: 0s - loss: 0.5402 - accuracy: 0.7400 -
true_positives: 408.0000 - true_negatives: 36.0000 - false_positives: 141.0000 - fals
e_negatives: 15.00 - 0s 3ms/step - loss: 0.5419 - accuracy: 0.7272 - true_positives:
5199.0000 - true_negatives: 374.0000 - false_positives: 1832.0000 - false_negatives:
259.0000 - val_loss: 0.5528 - val_accuracy: 0.7197 - val_true_positives: 1276.0000 -
val_true_negatives: 103.0000 - val_false_positives: 453.0000 - val_false_negatives: 8
4.0000
Epoch 7/100
13/13 [==============================] - ETA: 0s - loss: 0.5292 - accuracy: 0.7467 -
true_positives: 408.0000 - true_negatives: 40.0000 - false_positives: 130.0000 - fals
e_negatives: 22.00 - 0s 3ms/step - loss: 0.5327 - accuracy: 0.7300 - true_positives:
5069.0000 - true_negatives: 526.0000 - false_positives: 1680.0000 - false_negatives:
389.0000 - val_loss: 0.5461 - val_accuracy: 0.7171 - val_true_positives: 1253.0000 -
val_true_negatives: 121.0000 - val_false_positives: 435.0000 - val_false_negatives: 1
07.0000
Epoch 8/100
13/13 [==============================] - ETA: 0s - loss: 0.4941 - accuracy: 0.7583 -
true_positives: 417.0000 - true_negatives: 38.0000 - false_positives: 121.0000 - fals
e_negatives: 24.00 - 0s 3ms/step - loss: 0.5210 - accuracy: 0.7380 - true_positives:
```

```
5067.0000 - true_negatives: 589.0000 - false_positives: 1617.0000 - false_negatives:
391.0000 - val_loss: 0.5346 - val_accuracy: 0.7208 - val_true_positives: 1227.0000 -
val_true_negatives: 154.0000 - val_false_positives: 402.0000 - val_false_negatives: 1
33.0000
Epoch 9/100
13/13 [==============================] - ETA: 0s - loss: 0.5098 - accuracy: 0.7367 -
true_positives: 386.0000 - true_negatives: 56.0000 - false_positives: 126.0000 - fals
e_negatives: 32.00 - 0s 3ms/step - loss: 0.5079 - accuracy: 0.7450 - true_positives:
5097.0000 - true_negatives: 613.0000 - false_positives: 1593.0000 - false_negatives:
361.0000 - val_loss: 0.5232 - val_accuracy: 0.7307 - val_true_positives: 1251.0000 -
val_true_negatives: 149.0000 - val_false_positives: 407.0000 - val_false_negatives: 1
09.0000
Epoch 10/100
13/13 [==============================] - ETA: 0s - loss: 0.5219 - accuracy: 0.7450 -
true_positives: 390.0000 - true_negatives: 57.0000 - false_positives: 131.0000 - fals
e_negatives: 22.00 - 0s 2ms/step - loss: 0.4937 - accuracy: 0.7501 - true_positives:
5027.0000 - true_negatives: 722.0000 - false_positives: 1484.0000 - false_negatives:
431.0000 - val_loss: 0.5269 - val_accuracy: 0.7302 - val_true_positives: 1292.0000 -
val_true_negatives: 107.0000 - val_false_positives: 449.0000 - val_false_negatives: 6
8.0000
Epoch 11/100
13/13 [==============================] - ETA: 0s - loss: 0.5076 - accuracy: 0.7617 -
true_positives: 410.0000 - true_negatives: 47.0000 - false_positives: 132.0000 - fals
e_negatives: 11.00 - 0s 2ms/step - loss: 0.4871 - accuracy: 0.7560 - true_positives:
5004.0000 - true_negatives: 790.0000 - false_positives: 1416.0000 - false_negatives:
454.0000 - val_loss: 0.5114 - val_accuracy: 0.7474 - val_true_positives: 1279.0000 -
val_true_negatives: 153.0000 - val_false_positives: 403.0000 - val_false_negatives: 8
1.0000
Epoch 12/100
13/13 [==============================] - ETA: 0s - loss: 0.5368 - accuracy: 0.7383 -
true_positives: 400.0000 - true_negatives: 43.0000 - false_positives: 115.0000 - fals
e_negatives: 42.00 - 0s 3ms/step - loss: 0.4744 - accuracy: 0.7667 - true_positives:
4985.0000 - true_negatives: 891.0000 - false_positives: 1315.0000 - false_negatives:
473.0000 - val_loss: 0.5040 - val_accuracy: 0.7422 - val_true_positives: 1233.0000 -
val_true_negatives: 189.0000 - val_false_positives: 367.0000 - val_false_negatives: 1
27.0000
Epoch 13/100
13/13 [==============================] - ETA: 0s - loss: 0.4544 - accuracy: 0.7850 -
true_positives: 397.0000 - true_negatives: 74.0000 - false_positives: 100.0000 - fals
e_negatives: 29.00 - 0s 3ms/step - loss: 0.4654 - accuracy: 0.7706 - true_positives:
4954.0000 - true_negatives: 952.0000 - false_positives: 1254.0000 - false_negatives:
504.0000 - val_loss: 0.5117 - val_accuracy: 0.7479 - val_true_positives: 1276.0000 -
val_true_negatives: 157.0000 - val_false_positives: 399.0000 - val_false_negatives: 8
4.0000
Epoch 14/100
13/13 [==============================] - ETA: 0s - loss: 0.4948 - accuracy: 0.7783 -
true_positives: 401.0000 - true_negatives: 66.0000 - false_positives: 111.0000 - fals
e_negatives: 22.00 - 0s 2ms/step - loss: 0.4689 - accuracy: 0.7714 - true_positives:
4906.0000 - true_negatives: 1006.0000 - false_positives: 1200.0000 - false_negatives:
552.0000 - val_loss: 0.5160 - val_accuracy: 0.7448 - val_true_positives: 1283.0000 -
val_true_negatives: 144.0000 - val_false_positives: 412.0000 - val_false_negatives: 7
7.0000
Epoch 15/100
13/13 [==============================] - ETA: 0s - loss: 0.4590 - accuracy: 0.7717 -
true_positives: 407.0000 - true_negatives: 56.0000 - false_positives: 116.0000 - fals
e_negatives: 21.00 - 0s 2ms/step - loss: 0.4601 - accuracy: 0.7797 - true_positives:
4945.0000 - true_negatives: 1031.0000 - false_positives: 1175.0000 - false_negatives:
513.0000 - val_loss: 0.4988 - val_accuracy: 0.7516 - val_true_positives: 1172.0000 -
val_true_negatives: 268.0000 - val_false_positives: 288.0000 - val_false_negatives: 1
88.0000
```

```
Epoch 16/100
13/13 [==============================] - ETA: 0s - loss: 0.4540 - accuracy: 0.7850 -
true_positives: 371.0000 - true_negatives: 100.0000 - false_positives: 75.0000 - fals
e_negatives: 54.00 - 0s 2ms/step - loss: 0.4548 - accuracy: 0.7801 - true_positives:
4984.0000 - true_negatives: 995.0000 - false_positives: 1211.0000 - false_negatives:
474.0000 - val_loss: 0.4987 - val_accuracy: 0.7557 - val_true_positives: 1222.0000 -
val_true_negatives: 226.0000 - val_false_positives: 330.0000 - val_false_negatives: 1
38.0000
Epoch 17/100
13/13 [==============================] - ETA: 0s - loss: 0.4457 - accuracy: 0.7750 -
true_positives: 379.0000 - true_negatives: 86.0000 - false_positives: 83.0000 - false
_negatives: 52.000 - 0s 2ms/step - loss: 0.4484 - accuracy: 0.7837 - true_positives:
4888.0000 - true_negatives: 1118.0000 - false_positives: 1088.0000 - false_negatives:
570.0000 - val_loss: 0.4982 - val_accuracy: 0.7531 - val_true_positives: 1230.0000 -
val_true_negatives: 213.0000 - val_false_positives: 343.0000 - val_false_negatives: 1
30.0000
Epoch 18/100
13/13 [==============================] - ETA: 0s - loss: 0.4671 - accuracy: 0.7650 -
true_positives: 400.0000 - true_negatives: 59.0000 - false_positives: 100.0000 - fals
e_negatives: 41.00 - 0s 2ms/step - loss: 0.4443 - accuracy: 0.7878 - true_positives:
4953.0000 - true_negatives: 1085.0000 - false_positives: 1121.0000 - false_negatives:
505.0000 - val_loss: 0.5018 - val_accuracy: 0.7542 - val_true_positives: 1154.0000 -
val_true_negatives: 291.0000 - val_false_positives: 265.0000 - val_false_negatives: 2
06.0000
Epoch 19/100
13/13 [==============================] - ETA: 0s - loss: 0.4306 - accuracy: 0.7900 -
true_positives: 390.0000 - true_negatives: 84.0000 - false_positives: 59.0000 - false
_negatives: 67.000 - 0s 2ms/step - loss: 0.4478 - accuracy: 0.7851 - true_positives:
4916.0000 - true_negatives: 1101.0000 - false_positives: 1105.0000 - false_negatives:
542.0000 - val_loss: 0.5021 - val_accuracy: 0.7557 - val_true_positives: 1148.0000 -
val_true_negatives: 300.0000 - val_false_positives: 256.0000 - val_false_negatives: 2
12.0000
Epoch 20/100
13/13 [==============================] - ETA: 0s - loss: 0.4396 - accuracy: 0.7833 -
true_positives: 355.0000 - true_negatives: 115.0000 - false_positives: 72.0000 - fals
e_negatives: 58.00 - 0s 2ms/step - loss: 0.4428 - accuracy: 0.7891 - true_positives:
4920.0000 - true_negatives: 1128.0000 - false_positives: 1078.0000 - false_negatives:
538.0000 - val_loss: 0.5004 - val_accuracy: 0.7521 - val_true_positives: 1202.0000 -
val_true_negatives: 239.0000 - val_false_positives: 317.0000 - val_false_negatives: 1
58.0000
Epoch 21/100
13/13 [==============================] - ETA: 0s - loss: 0.4646 - accuracy: 0.7683 -
true_positives: 376.0000 - true_negatives: 85.0000 - false_positives: 90.0000 - false
_negatives: 49.000 - 0s 2ms/step - loss: 0.4420 - accuracy: 0.7888 - true_positives:
4881.0000 - true_negatives: 1164.0000 - false_positives: 1042.0000 - false_negatives:
577.0000 - val_loss: 0.5135 - val_accuracy: 0.7542 - val_true_positives: 1277.0000 -
val_true_negatives: 168.0000 - val_false_positives: 388.0000 - val_false_negatives: 8
3.0000
Epoch 22/100
13/13 [==============================] - ETA: 0s - loss: 0.4023 - accuracy: 0.8217 -
true_positives: 420.0000 - true_negatives: 73.0000 - false_positives: 92.0000 - false
_negatives: 15.000 - 0s 2ms/step - loss: 0.4353 - accuracy: 0.7968 - true_positives:
4981.0000 - true_negatives: 1126.0000 - false_positives: 1080.0000 - false_negatives:
477.0000 - val_loss: 0.4994 - val_accuracy: 0.7568 - val_true_positives: 1167.0000 -
val_true_negatives: 283.0000 - val_false_positives: 273.0000 - val_false_negatives: 1
93.0000
Epoch 23/100
13/13 [==============================] - ETA: 0s - loss: 0.4106 - accuracy: 0.8033 -
true_positives: 386.0000 - true_negatives: 96.0000 - false_positives: 71.0000 - false
_negatives: 47.000 - 0s 2ms/step - loss: 0.4319 - accuracy: 0.7958 - true_positives:
```

```
4945.0000 - true_negatives: 1154.0000 - false_positives: 1052.0000 - false_negatives:
513.0000 - val_loss: 0.5044 - val_accuracy: 0.7557 - val_true_positives: 1231.0000 -
val_true_negatives: 217.0000 - val_false_positives: 339.0000 - val_false_negatives: 1
29.0000
Epoch 24/100
13/13 [==============================] - ETA: 0s - loss: 0.4551 - accuracy: 0.7850 -
true_positives: 386.0000 - true_negatives: 85.0000 - false_positives: 99.0000 - false
_negatives: 30.000 - 0s 2ms/step - loss: 0.4294 - accuracy: 0.8001 - true_positives:
4921.0000 - true_negatives: 1211.0000 - false_positives: 995.0000 - false_negatives:
537.0000 - val_loss: 0.5014 - val_accuracy: 0.7630 - val_true_positives: 1225.0000 -
val_true_negatives: 237.0000 - val_false_positives: 319.0000 - val_false_negatives: 1
35.0000
Epoch 25/100
13/13 [==============================] - ETA: 0s - loss: 0.4399 - accuracy: 0.7733 -
true_positives: 382.0000 - true_negatives: 82.0000 - false_positives: 92.0000 - false
_negatives: 44.000 - 0s 2ms/step - loss: 0.4282 - accuracy: 0.7978 - true_positives:
4967.0000 - true_negatives: 1147.0000 - false_positives: 1059.0000 - false_negatives:
491.0000 - val_loss: 0.5053 - val_accuracy: 0.7531 - val_true_positives: 1161.0000 -
val_true_negatives: 282.0000 - val_false_positives: 274.0000 - val_false_negatives: 1
99.0000
Epoch 26/100
13/13 [==============================] - ETA: 0s - loss: 0.3859 - accuracy: 0.8350 -
true_positives: 386.0000 - true_negatives: 115.0000 - false_positives: 55.0000 - fals
e_negatives: 44.00 - 0s 2ms/step - loss: 0.4262 - accuracy: 0.8022 - true_positives:
4941.0000 - true_negatives: 1207.0000 - false_positives: 999.0000 - false_negatives:
517.0000 - val_loss: 0.5041 - val_accuracy: 0.7599 - val_true_positives: 1212.0000 -
val_true_negatives: 244.0000 - val_false_positives: 312.0000 - val_false_negatives: 1
48.0000
Epoch 27/100
13/13 [==============================] - ETA: 0s - loss: 0.4155 - accuracy: 0.8033 -
true_positives: 384.0000 - true_negatives: 98.0000 - false_positives: 84.0000 - false
_negatives: 34.000 - 0s 2ms/step - loss: 0.4276 - accuracy: 0.7987 - true_positives:
4890.0000 - true_negatives: 1231.0000 - false_positives: 975.0000 - false_negatives:
568.0000 - val_loss: 0.5267 - val_accuracy: 0.7495 - val_true_positives: 1286.0000 -
val_true_negatives: 150.0000 - val_false_positives: 406.0000 - val_false_negatives: 7
4.0000
Epoch 28/100
13/13 [==============================] - ETA: 0s - loss: 0.4209 - accuracy: 0.8250 -
true_positives: 415.0000 - true_negatives: 80.0000 - false_positives: 94.0000 - false
_negatives: 11.000 - 0s 2ms/step - loss: 0.4217 - accuracy: 0.8038 - true_positives:
4951.0000 - true_negatives: 1209.0000 - false_positives: 997.0000 - false_negatives:
507.0000 - val_loss: 0.5069 - val_accuracy: 0.7521 - val_true_positives: 1168.0000 -
val_true_negatives: 273.0000 - val_false_positives: 283.0000 - val_false_negatives: 1
92.0000
Epoch 29/100
13/13 [==============================] - ETA: 0s - loss: 0.4174 - accuracy: 0.8150 -
true_positives: 383.0000 - true_negatives: 106.0000 - false_positives: 77.0000 - fals
e_negatives: 34.00 - 0s 2ms/step - loss: 0.4176 - accuracy: 0.8072 - true_positives:
4946.0000 - true_negatives: 1240.0000 - false_positives: 966.0000 - false_negatives:
512.0000 - val_loss: 0.5110 - val_accuracy: 0.7589 - val_true_positives: 1228.0000 -
val_true_negatives: 226.0000 - val_false_positives: 330.0000 - val_false_negatives: 1
32.0000
Epoch 30/100
13/13 [==============================] - ETA: 0s - loss: 0.3701 - accuracy: 0.8333 -
true_positives: 406.0000 - true_negatives: 94.0000 - false_positives: 73.0000 - false
_negatives: 27.000 - 0s 2ms/step - loss: 0.4163 - accuracy: 0.8072 - true_positives:
4977.0000 - true_negatives: 1209.0000 - false_positives: 997.0000 - false_negatives:
481.0000 - val_loss: 0.5090 - val_accuracy: 0.7563 - val_true_positives: 1170.0000 -
val_true_negatives: 279.0000 - val_false_positives: 277.0000 - val_false_negatives: 1
90.0000
```

```
Epoch 31/100
13/13 [==============================] - ETA: 0s - loss: 0.3855 - accuracy: 0.8283 -
true_positives: 401.0000 - true_negatives: 96.0000 - false_positives: 66.0000 - false
_negatives: 37.000 - 0s 2ms/step - loss: 0.4110 - accuracy: 0.8078 - true_positives:
4970.0000 - true_negatives: 1221.0000 - false_positives: 985.0000 - false_negatives:
488.0000 - val_loss: 0.5112 - val_accuracy: 0.7542 - val_true_positives: 1205.0000 -
val_true_negatives: 240.0000 - val_false_positives: 316.0000 - val_false_negatives: 1
55.0000
Epoch 32/100
13/13 [==============================] - ETA: 0s - loss: 0.4216 - accuracy: 0.8083 -
true_positives: 404.0000 - true_negatives: 81.0000 - false_positives: 77.0000 - false
_negatives: 38.000 - 0s 2ms/step - loss: 0.4107 - accuracy: 0.8091 - true_positives:
4946.0000 - true_negatives: 1255.0000 - false_positives: 951.0000 - false_negatives:
512.0000 - val_loss: 0.5131 - val_accuracy: 0.7542 - val_true_positives: 1181.0000 -
val_true_negatives: 264.0000 - val_false_positives: 292.0000 - val_false_negatives: 1
79.0000
Epoch 33/100
13/13 [==============================] - ETA: 0s - loss: 0.4125 - accuracy: 0.8133 -
true_positives: 391.0000 - true_negatives: 97.0000 - false_positives: 71.0000 - false
_negatives: 41.000 - 0s 2ms/step - loss: 0.4128 - accuracy: 0.8098 - true_positives:
4942.0000 - true_negatives: 1264.0000 - false_positives: 942.0000 - false_negatives:
516.0000 - val_loss: 0.5170 - val_accuracy: 0.7526 - val_true_positives: 1155.0000 -
val_true_negatives: 287.0000 - val_false_positives: 269.0000 - val_false_negatives: 2
05.0000
Epoch 34/100
13/13 [==============================] - ETA: 0s - loss: 0.4144 - accuracy: 0.8067 -
true_positives: 375.0000 - true_negatives: 109.0000 - false_positives: 68.0000 - fals
e_negatives: 48.00 - 0s 2ms/step - loss: 0.4097 - accuracy: 0.8120 - true_positives:
4989.0000 - true_negatives: 1234.0000 - false_positives: 972.0000 - false_negatives:
469.0000 - val_loss: 0.5184 - val_accuracy: 0.7526 - val_true_positives: 1168.0000 -
val_true_negatives: 274.0000 - val_false_positives: 282.0000 - val_false_negatives: 1
92.0000
Epoch 35/100
13/13 [==============================] - ETA: 0s - loss: 0.4040 - accuracy: 0.8100 -
true_positives: 380.0000 - true_negatives: 106.0000 - false_positives: 61.0000 - fals
e_negatives: 53.00 - 0s 2ms/step - loss: 0.4083 - accuracy: 0.8126 - true_positives:
4969.0000 - true_negatives: 1259.0000 - false_positives: 947.0000 - false_negatives:
489.0000 - val_loss: 0.5203 - val_accuracy: 0.7469 - val_true_positives: 1144.0000 -
val_true_negatives: 287.0000 - val_false_positives: 269.0000 - val_false_negatives: 2
16.0000
Epoch 36/100
13/13 [==============================] - ETA: 0s - loss: 0.3913 - accuracy: 0.8183 -
true_positives: 378.0000 - true_negatives: 113.0000 - false_positives: 55.0000 - fals
e_negatives: 54.00 - 0s 2ms/step - loss: 0.4023 - accuracy: 0.8143 - true_positives:
4959.0000 - true_negatives: 1282.0000 - false_positives: 924.0000 - false_negatives:
499.0000 - val_loss: 0.5167 - val_accuracy: 0.7604 - val_true_positives: 1218.0000 -
val_true_negatives: 239.0000 - val_false_positives: 317.0000 - val_false_negatives: 1
42.0000
Epoch 37/100
13/13 [==============================] - ETA: 0s - loss: 0.3989 - accuracy: 0.8100 -
true_positives: 381.0000 - true_negatives: 105.0000 - false_positives: 80.0000 - fals
e_negatives: 34.00 - 0s 2ms/step - loss: 0.3988 - accuracy: 0.8176 - true_positives:
4985.0000 - true_negatives: 1281.0000 - false_positives: 925.0000 - false_negatives:
473.0000 - val_loss: 0.5229 - val_accuracy: 0.7500 - val_true_positives: 1173.0000 -
val_true_negatives: 264.0000 - val_false_positives: 292.0000 - val_false_negatives: 1
87.0000
Epoch 38/100
13/13 [==============================] - ETA: 0s - loss: 0.3714 - accuracy: 0.8400 -
true_positives: 399.0000 - true_negatives: 105.0000 - false_positives: 62.0000 - fals
e_negatives: 34.00 - 0s 2ms/step - loss: 0.3979 - accuracy: 0.8210 - true_positives:
```

```
5010.0000 - true_negatives: 1282.0000 - false_positives: 924.0000 - false_negatives:
448.0000 - val_loss: 0.5209 - val_accuracy: 0.7557 - val_true_positives: 1185.0000 -
val_true_negatives: 263.0000 - val_false_positives: 293.0000 - val_false_negatives: 1
75.0000
Epoch 39/100
13/13 [==============================] - ETA: 0s - loss: 0.4047 - accuracy: 0.8117 -
true_positives: 403.0000 - true_negatives: 84.0000 - false_positives: 71.0000 - false
_negatives: 42.000 - 0s 2ms/step - loss: 0.3975 - accuracy: 0.8175 - true_positives:
4951.0000 - true_negatives: 1314.0000 - false_positives: 892.0000 - false_negatives:
507.0000 - val_loss: 0.5300 - val_accuracy: 0.7458 - val_true_positives: 1132.0000 -
val_true_negatives: 297.0000 - val_false_positives: 259.0000 - val_false_negatives: 2
28.0000
Epoch 40/100
13/13 [==============================] - ETA: 0s - loss: 0.4090 - accuracy: 0.8150 -
true_positives: 345.0000 - true_negatives: 144.0000 - false_positives: 65.0000 - fals
e_negatives: 46.00 - 0s 2ms/step - loss: 0.4010 - accuracy: 0.8168 - true_positives:
4970.0000 - true_negatives: 1290.0000 - false_positives: 916.0000 - false_negatives:
488.0000 - val_loss: 0.5241 - val_accuracy: 0.7490 - val_true_positives: 1143.0000 -
val_true_negatives: 292.0000 - val_false_positives: 264.0000 - val_false_negatives: 2
17.0000
Epoch 41/100
13/13 [==============================] - ETA: 0s - loss: 0.3684 - accuracy: 0.8400 -
true_positives: 384.0000 - true_negatives: 120.0000 - false_positives: 50.0000 - fals
e_negatives: 46.00 - 0s 2ms/step - loss: 0.3916 - accuracy: 0.8220 - true_positives:
4991.0000 - true_negatives: 1309.0000 - false_positives: 897.0000 - false_negatives:
467.0000 - val_loss: 0.5281 - val_accuracy: 0.7510 - val_true_positives: 1175.0000 -
val_true_negatives: 264.0000 - val_false_positives: 292.0000 - val_false_negatives: 1
85.0000
Epoch 42/100
13/13 [==============================] - ETA: 0s - loss: 0.3832 - accuracy: 0.8367 -
true_positives: 405.0000 - true_negatives: 97.0000 - false_positives: 54.0000 - false
_negatives: 44.000 - 0s 2ms/step - loss: 0.3945 - accuracy: 0.8173 - true_positives:
4941.0000 - true_negatives: 1323.0000 - false_positives: 883.0000 - false_negatives:
517.0000 - val_loss: 0.5413 - val_accuracy: 0.7542 - val_true_positives: 1260.0000 -
val_true_negatives: 185.0000 - val_false_positives: 371.0000 - val_false_negatives: 1
00.0000
Epoch 43/100
13/13 [==============================] - ETA: 0s - loss: 0.4349 - accuracy: 0.8033 -
true_positives: 394.0000 - true_negatives: 88.0000 - false_positives: 108.0000 - fals
e_negatives: 10.00 - 0s 2ms/step - loss: 0.3942 - accuracy: 0.8159 - true_positives:
4988.0000 - true_negatives: 1265.0000 - false_positives: 941.0000 - false_negatives:
470.0000 - val_loss: 0.5345 - val_accuracy: 0.7500 - val_true_positives: 1221.0000 -
val_true_negatives: 216.0000 - val_false_positives: 340.0000 - val_false_negatives: 1
39.0000
Epoch 44/100
13/13 [==============================] - ETA: 0s - loss: 0.3766 - accuracy: 0.8233 -
true_positives: 397.0000 - true_negatives: 97.0000 - false_positives: 73.0000 - false
_negatives: 33.000 - 0s 2ms/step - loss: 0.3870 - accuracy: 0.8220 - true_positives:
4962.0000 - true_negatives: 1338.0000 - false_positives: 868.0000 - false_negatives:
496.0000 - val_loss: 0.5338 - val_accuracy: 0.7552 - val_true_positives: 1232.0000 -
val_true_negatives: 215.0000 - val_false_positives: 341.0000 - val_false_negatives: 1
28.0000
Epoch 45/100
13/13 [==============================] - ETA: 0s - loss: 0.3804 - accuracy: 0.8283 -
true_positives: 397.0000 - true_negatives: 100.0000 - false_positives: 82.0000 - fals
e_negatives: 21.00 - 0s 2ms/step - loss: 0.3826 - accuracy: 0.8214 - true_positives:
4996.0000 - true_negatives: 1299.0000 - false_positives: 907.0000 - false_negatives:
462.0000 - val_loss: 0.5349 - val_accuracy: 0.7521 - val_true_positives: 1209.0000 -
val_true_negatives: 232.0000 - val_false_positives: 324.0000 - val_false_negatives: 1
51.0000
```

```
Epoch 46/100
13/13 [==============================] - ETA: 0s - loss: 0.3877 - accuracy: 0.8333 -
true_positives: 396.0000 - true_negatives: 104.0000 - false_positives: 74.0000 - fals
e_negatives: 26.00 - 0s 2ms/step - loss: 0.3776 - accuracy: 0.8291 - true_positives:
4989.0000 - true_negatives: 1365.0000 - false_positives: 841.0000 - false_negatives:
469.0000 - val_loss: 0.5374 - val_accuracy: 0.7557 - val_true_positives: 1203.0000 -
val_true_negatives: 245.0000 - val_false_positives: 311.0000 - val_false_negatives: 1
57.0000
Epoch 47/100
13/13 [==============================] - ETA: 0s - loss: 0.3829 - accuracy: 0.8183 -
true_positives: 381.0000 - true_negatives: 110.0000 - false_positives: 66.0000 - fals
e_negatives: 43.00 - 0s 2ms/step - loss: 0.3782 - accuracy: 0.8289 - true_positives:
4979.0000 - true_negatives: 1374.0000 - false_positives: 832.0000 - false_negatives:
479.0000 - val_loss: 0.5450 - val_accuracy: 0.7573 - val_true_positives: 1238.0000 -
val_true_negatives: 213.0000 - val_false_positives: 343.0000 - val_false_negatives: 1
22.0000
Epoch 48/100
13/13 [==============================] - ETA: 0s - loss: 0.3798 - accuracy: 0.8333 -
true_positives: 416.0000 - true_negatives: 84.0000 - false_positives: 84.0000 - false
_negatives: 16.000 - 0s 2ms/step - loss: 0.3806 - accuracy: 0.8244 - true_positives:
5003.0000 - true_negatives: 1315.0000 - false_positives: 891.0000 - false_negatives:
455.0000 - val_loss: 0.5387 - val_accuracy: 0.7557 - val_true_positives: 1207.0000 -
val_true_negatives: 241.0000 - val_false_positives: 315.0000 - val_false_negatives: 1
53.0000
Epoch 49/100
13/13 [==============================] - ETA: 0s - loss: 0.3982 - accuracy: 0.8233 -
true_positives: 398.0000 - true_negatives: 96.0000 - false_positives: 74.0000 - false
_negatives: 32.000 - 0s 2ms/step - loss: 0.3755 - accuracy: 0.8287 - true_positives:
4995.0000 - true_negatives: 1356.0000 - false_positives: 850.0000 - false_negatives:
463.0000 - val_loss: 0.5416 - val_accuracy: 0.7531 - val_true_positives: 1197.0000 -
val_true_negatives: 246.0000 - val_false_positives: 310.0000 - val_false_negatives: 1
63.0000
Epoch 50/100
13/13 [==============================] - ETA: 0s - loss: 0.3603 - accuracy: 0.8500 -
true_positives: 404.0000 - true_negatives: 106.0000 - false_positives: 62.0000 - fals
e_negatives: 28.00 - 0s 2ms/step - loss: 0.3727 - accuracy: 0.8330 - true_positives:
4995.0000 - true_negatives: 1389.0000 - false_positives: 817.0000 - false_negatives:
463.0000 - val_loss: 0.5594 - val_accuracy: 0.7495 - val_true_positives: 1252.0000 -
val_true_negatives: 184.0000 - val_false_positives: 372.0000 - val_false_negatives: 1
08.0000
Epoch 51/100
13/13 [==============================] - ETA: 0s - loss: 0.3751 - accuracy: 0.8317 -
true_positives: 407.0000 - true_negatives: 92.0000 - false_positives: 88.0000 - false
_negatives: 13.000 - 0s 2ms/step - loss: 0.3819 - accuracy: 0.8245 - true_positives:
4995.0000 - true_negatives: 1324.0000 - false_positives: 882.0000 - false_negatives:
463.0000 - val_loss: 0.5473 - val_accuracy: 0.7516 - val_true_positives: 1212.0000 -
val_true_negatives: 228.0000 - val_false_positives: 328.0000 - val_false_negatives: 1
48.0000
Epoch 52/100
13/13 [==============================] - ETA: 0s - loss: 0.3765 - accuracy: 0.8250 -
true_positives: 410.0000 - true_negatives: 85.0000 - false_positives: 75.0000 - false
_negatives: 30.000 - 0s 3ms/step - loss: 0.3730 - accuracy: 0.8284 - true_positives:
4966.0000 - true_negatives: 1383.0000 - false_positives: 823.0000 - false_negatives:
492.0000 - val_loss: 0.5546 - val_accuracy: 0.7521 - val_true_positives: 1219.0000 -
val_true_negatives: 222.0000 - val_false_positives: 334.0000 - val_false_negatives: 1
41.0000
Epoch 53/100
13/13 [==============================] - ETA: 0s - loss: 0.3802 - accuracy: 0.8300 -
true_positives: 411.0000 - true_negatives: 87.0000 - false_positives: 73.0000 - false
_negatives: 29.000 - 0s 3ms/step - loss: 0.3717 - accuracy: 0.8317 - true_positives:
```

```
5005.0000 - true_negatives: 1369.0000 - false_positives: 837.0000 - false_negatives:
453.0000 - val_loss: 0.5542 - val_accuracy: 0.7469 - val_true_positives: 1196.0000 -
val_true_negatives: 235.0000 - val_false_positives: 321.0000 - val_false_negatives: 1
64.0000
Epoch 54/100
13/13 [==============================] - ETA: 0s - loss: 0.3400 - accuracy: 0.8483 -
true_positives: 402.0000 - true_negatives: 107.0000 - false_positives: 67.0000 - fals
e_negatives: 24.00 - 0s 2ms/step - loss: 0.3668 - accuracy: 0.8347 - true_positives:
5025.0000 - true_negatives: 1372.0000 - false_positives: 834.0000 - false_negatives:
433.0000 - val_loss: 0.5630 - val_accuracy: 0.7396 - val_true_positives: 1127.0000 -
val_true_negatives: 290.0000 - val_false_positives: 266.0000 - val_false_negatives: 2
33.0000
Epoch 55/100
13/13 [==============================] - ETA: 0s - loss: 0.3875 - accuracy: 0.8283 -
true_positives: 365.0000 - true_negatives: 132.0000 - false_positives: 46.0000 - fals
e_negatives: 57.00 - 0s 2ms/step - loss: 0.3753 - accuracy: 0.8288 - true_positives:
4976.0000 - true_negatives: 1376.0000 - false_positives: 830.0000 - false_negatives:
482.0000 - val_loss: 0.5538 - val_accuracy: 0.7578 - val_true_positives: 1201.0000 -
val_true_negatives: 251.0000 - val_false_positives: 305.0000 - val_false_negatives: 1
59.0000
Epoch 56/100
13/13 [==============================] - ETA: 0s - loss: 0.3671 - accuracy: 0.8233 -
true_positives: 390.0000 - true_negatives: 104.0000 - false_positives: 71.0000 - fals
e_negatives: 35.00 - 0s 2ms/step - loss: 0.3698 - accuracy: 0.8304 - true_positives:
4995.0000 - true_negatives: 1369.0000 - false_positives: 837.0000 - false_negatives:
463.0000 - val_loss: 0.5515 - val_accuracy: 0.7469 - val_true_positives: 1197.0000 -
val_true_negatives: 234.0000 - val_false_positives: 322.0000 - val_false_negatives: 1
63.0000
Epoch 57/100
13/13 [==============================] - ETA: 0s - loss: 0.3420 - accuracy: 0.8533 -
true_positives: 409.0000 - true_negatives: 103.0000 - false_positives: 59.0000 - fals
e_negatives: 29.00 - 0s 2ms/step - loss: 0.3663 - accuracy: 0.8327 - true_positives:
5011.0000 - true_negatives: 1371.0000 - false_positives: 835.0000 - false_negatives:
447.0000 - val_loss: 0.5579 - val_accuracy: 0.7516 - val_true_positives: 1180.0000 -
val_true_negatives: 260.0000 - val_false_positives: 296.0000 - val_false_negatives: 1
80.0000
Epoch 58/100
13/13 [==============================] - ETA: 0s - loss: 0.3786 - accuracy: 0.8367 -
true_positives: 384.0000 - true_negatives: 118.0000 - false_positives: 61.0000 - fals
e_negatives: 37.00 - 0s 2ms/step - loss: 0.3612 - accuracy: 0.8376 - true_positives:
5038.0000 - true_negatives: 1381.0000 - false_positives: 825.0000 - false_negatives:
420.0000 - val_loss: 0.5625 - val_accuracy: 0.7443 - val_true_positives: 1168.0000 -
val_true_negatives: 258.0000 - val_false_positives: 298.0000 - val_false_negatives: 1
92.0000
Epoch 59/100
13/13 [==============================] - ETA: 0s - loss: 0.3891 - accuracy: 0.8417 -
true_positives: 384.0000 - true_negatives: 121.0000 - false_positives: 58.0000 - fals
e_negatives: 37.00 - 0s 2ms/step - loss: 0.3612 - accuracy: 0.8370 - true_positives:
5015.0000 - true_negatives: 1400.0000 - false_positives: 806.0000 - false_negatives:
443.0000 - val_loss: 0.5656 - val_accuracy: 0.7484 - val_true_positives: 1170.0000 -
val_true_negatives: 264.0000 - val_false_positives: 292.0000 - val_false_negatives: 1
90.0000
Epoch 60/100
13/13 [==============================] - ETA: 0s - loss: 0.3408 - accuracy: 0.8633 -
true_positives: 410.0000 - true_negatives: 108.0000 - false_positives: 48.0000 - fals
e_negatives: 34.00 - 0s 2ms/step - loss: 0.3606 - accuracy: 0.8378 - true_positives:
5025.0000 - true_negatives: 1396.0000 - false_positives: 810.0000 - false_negatives:
433.0000 - val_loss: 0.5651 - val_accuracy: 0.7505 - val_true_positives: 1172.0000 -
val_true_negatives: 266.0000 - val_false_positives: 290.0000 - val_false_negatives: 1
88.0000
```

```
Epoch 61/100
13/13 [==============================] - ETA: 0s - loss: 0.3514 - accuracy: 0.8400 -
true_positives: 403.0000 - true_negatives: 101.0000 - false_positives: 59.0000 - fals
e_negatives: 37.00 - 0s 2ms/step - loss: 0.3570 - accuracy: 0.8381 - true_positives:
5008.0000 - true_negatives: 1415.0000 - false_positives: 791.0000 - false_negatives:
450.0000 - val_loss: 0.5706 - val_accuracy: 0.7469 - val_true_positives: 1204.0000 -
val_true_negatives: 227.0000 - val_false_positives: 329.0000 - val_false_negatives: 1
56.0000
Epoch 62/100
13/13 [==============================] - ETA: 0s - loss: 0.3246 - accuracy: 0.8433 -
true_positives: 393.0000 - true_negatives: 113.0000 - false_positives: 67.0000 - fals
e_negatives: 27.00 - 0s 2ms/step - loss: 0.3574 - accuracy: 0.8374 - true_positives:
5036.0000 - true_negatives: 1382.0000 - false_positives: 824.0000 - false_negatives:
422.0000 - val_loss: 0.5693 - val_accuracy: 0.7448 - val_true_positives: 1196.0000 -
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
64.0000
Epoch 63/100
13/13 [==============================] - ETA: 0s - loss: 0.3248 - accuracy: 0.8467 -
true_positives: 412.0000 - true_negatives: 96.0000 - false_positives: 71.0000 - false
_negatives: 21.000 - 0s 3ms/step - loss: 0.3529 - accuracy: 0.8366 - true_positives:
5019.0000 - true_negatives: 1393.0000 - false_positives: 813.0000 - false_negatives:
439.0000 - val_loss: 0.5762 - val_accuracy: 0.7411 - val_true_positives: 1204.0000 -
val_true_negatives: 216.0000 - val_false_positives: 340.0000 - val_false_negatives: 1
56.0000
Epoch 64/100
13/13 [==============================] - ETA: 0s - loss: 0.3454 - accuracy: 0.8400 -
true_positives: 412.0000 - true_negatives: 92.0000 - false_positives: 65.0000 - false
_negatives: 31.000 - 0s 2ms/step - loss: 0.3543 - accuracy: 0.8368 - true_positives:
4991.0000 - true_negatives: 1422.0000 - false_positives: 784.0000 - false_negatives:
467.0000 - val_loss: 0.5876 - val_accuracy: 0.7453 - val_true_positives: 1237.0000 -
val_true_negatives: 191.0000 - val_false_positives: 365.0000 - val_false_negatives: 1
23.0000
Epoch 65/100
13/13 [==============================] - ETA: 0s - loss: 0.3087 - accuracy: 0.8483 -
true_positives: 408.0000 - true_negatives: 101.0000 - false_positives: 73.0000 - fals
e_negatives: 18.00 - 0s 2ms/step - loss: 0.3526 - accuracy: 0.8376 - true_positives:
5025.0000 - true_negatives: 1394.0000 - false_positives: 812.0000 - false_negatives:
433.0000 - val_loss: 0.5856 - val_accuracy: 0.7380 - val_true_positives: 1142.0000 -
val_true_negatives: 272.0000 - val_false_positives: 284.0000 - val_false_negatives: 2
18.0000
Epoch 66/100
13/13 [==============================] - ETA: 0s - loss: 0.3126 - accuracy: 0.8600 -
true_positives: 401.0000 - true_negatives: 115.0000 - false_positives: 43.0000 - fals
e_negatives: 41.00 - 0s 2ms/step - loss: 0.3518 - accuracy: 0.8379 - true_positives:
5009.0000 - true_negatives: 1413.0000 - false_positives: 793.0000 - false_negatives:
449.0000 - val_loss: 0.5793 - val_accuracy: 0.7484 - val_true_positives: 1161.0000 -
val_true_negatives: 273.0000 - val_false_positives: 283.0000 - val_false_negatives: 1
99.0000
Epoch 67/100
13/13 [==============================] - ETA: 0s - loss: 0.3647 - accuracy: 0.8333 -
true_positives: 385.0000 - true_negatives: 115.0000 - false_positives: 54.0000 - fals
e_negatives: 46.00 - 0s 2ms/step - loss: 0.3497 - accuracy: 0.8386 - true_positives:
5010.0000 - true_negatives: 1417.0000 - false_positives: 789.0000 - false_negatives:
448.0000 - val_loss: 0.5835 - val_accuracy: 0.7458 - val_true_positives: 1210.0000 -
val_true_negatives: 219.0000 - val_false_positives: 337.0000 - val_false_negatives: 1
50.0000
Epoch 68/100
13/13 [==============================] - ETA: 0s - loss: 0.3516 - accuracy: 0.8400 -
true_positives: 402.0000 - true_negatives: 102.0000 - false_positives: 71.0000 - fals
e_negatives: 25.00 - 0s 2ms/step - loss: 0.3486 - accuracy: 0.8424 - true_positives:
```

```
5018.0000 - true_negatives: 1438.0000 - false_positives: 768.0000 - false_negatives:
440.0000 - val_loss: 0.6173 - val_accuracy: 0.7416 - val_true_positives: 1238.0000 -
val_true_negatives: 183.0000 - val_false_positives: 373.0000 - val_false_negatives: 1
22.0000
Epoch 69/100
13/13 [==============================] - ETA: 0s - loss: 0.3510 - accuracy: 0.8383 -
true_positives: 405.0000 - true_negatives: 98.0000 - false_positives: 80.0000 - false
_negatives: 17.000 - 0s 2ms/step - loss: 0.3535 - accuracy: 0.8361 - true_positives:
5014.0000 - true_negatives: 1394.0000 - false_positives: 812.0000 - false_negatives:
444.0000 - val_loss: 0.5894 - val_accuracy: 0.7453 - val_true_positives: 1238.0000 -
val_true_negatives: 190.0000 - val_false_positives: 366.0000 - val_false_negatives: 1
22.0000
Epoch 70/100
13/13 [==============================] - ETA: 0s - loss: 0.3631 - accuracy: 0.8400 -
true_positives: 414.0000 - true_negatives: 90.0000 - false_positives: 82.0000 - false
_negatives: 14.000 - 0s 2ms/step - loss: 0.3484 - accuracy: 0.8409 - true_positives:
5024.0000 - true_negatives: 1421.0000 - false_positives: 785.0000 - false_negatives:
434.0000 - val_loss: 0.5880 - val_accuracy: 0.7411 - val_true_positives: 1186.0000 -
val_true_negatives: 234.0000 - val_false_positives: 322.0000 - val_false_negatives: 1
74.0000
Epoch 71/100
13/13 [==============================] - ETA: 0s - loss: 0.3090 - accuracy: 0.8617 -
true_positives: 415.0000 - true_negatives: 102.0000 - false_positives: 58.0000 - fals
e_negatives: 25.00 - 0s 2ms/step - loss: 0.3428 - accuracy: 0.8426 - true_positives:
5024.0000 - true_negatives: 1434.0000 - false_positives: 772.0000 - false_negatives:
434.0000 - val_loss: 0.5928 - val_accuracy: 0.7427 - val_true_positives: 1140.0000 -
val_true_negatives: 283.0000 - val_false_positives: 273.0000 - val_false_negatives: 2
20.0000
Epoch 72/100
13/13 [==============================] - ETA: 0s - loss: 0.3662 - accuracy: 0.8267 -
true_positives: 386.0000 - true_negatives: 110.0000 - false_positives: 55.0000 - fals
e_negatives: 49.00 - 0s 2ms/step - loss: 0.3544 - accuracy: 0.8359 - true_positives:
4992.0000 - true_negatives: 1414.0000 - false_positives: 792.0000 - false_negatives:
466.0000 - val_loss: 0.5901 - val_accuracy: 0.7390 - val_true_positives: 1157.0000 -
val_true_negatives: 259.0000 - val_false_positives: 297.0000 - val_false_negatives: 2
03.0000
Epoch 73/100
13/13 [==============================] - ETA: 0s - loss: 0.3049 - accuracy: 0.8750 -
true_positives: 417.0000 - true_negatives: 108.0000 - false_positives: 46.0000 - fals
e_negatives: 29.00 - 0s 2ms/step - loss: 0.3458 - accuracy: 0.8387 - true_positives:
5005.0000 - true_negatives: 1423.0000 - false_positives: 783.0000 - false_negatives:
453.0000 - val_loss: 0.5952 - val_accuracy: 0.7510 - val_true_positives: 1206.0000 -
val_true_negatives: 233.0000 - val_false_positives: 323.0000 - val_false_negatives: 1
54.0000
Epoch 74/100
13/13 [==============================] - ETA: 0s - loss: 0.3165 - accuracy: 0.8667 -
true_positives: 412.0000 - true_negatives: 108.0000 - false_positives: 60.0000 - fals
e_negatives: 20.00 - 0s 1ms/step - loss: 0.3414 - accuracy: 0.8447 - true_positives:
5064.0000 - true_negatives: 1410.0000 - false_positives: 796.0000 - false_negatives:
394.0000 - val_loss: 0.5987 - val_accuracy: 0.7422 - val_true_positives: 1180.0000 -
val_true_negatives: 242.0000 - val_false_positives: 314.0000 - val_false_negatives: 1
80.0000
Epoch 75/100
13/13 [==============================] - ETA: 0s - loss: 0.2835 - accuracy: 0.8767 -
true_positives: 405.0000 - true_negatives: 121.0000 - false_positives: 47.0000 - fals
e_negatives: 27.00 - 0s 1ms/step - loss: 0.3359 - accuracy: 0.8471 - true_positives:
5020.0000 - true_negatives: 1472.0000 - false_positives: 734.0000 - false_negatives:
438.0000 - val_loss: 0.6021 - val_accuracy: 0.7437 - val_true_positives: 1211.0000 -
val_true_negatives: 214.0000 - val_false_positives: 342.0000 - val_false_negatives: 1
49.0000
```

```
Epoch 76/100
13/13 [==============================] - ETA: 0s - loss: 0.3049 - accuracy: 0.8700 -
true_positives: 401.0000 - true_negatives: 121.0000 - false_positives: 55.0000 - fals
e_negatives: 23.00 - 0s 3ms/step - loss: 0.3401 - accuracy: 0.8460 - true_positives:
5042.0000 - true_negatives: 1442.0000 - false_positives: 764.0000 - false_negatives:
416.0000 - val_loss: 0.6168 - val_accuracy: 0.7448 - val_true_positives: 1220.0000 -
val_true_negatives: 207.0000 - val_false_positives: 349.0000 - val_false_negatives: 1
40.0000
Epoch 77/100
13/13 [==============================] - ETA: 0s - loss: 0.3214 - accuracy: 0.8533 -
true_positives: 405.0000 - true_negatives: 107.0000 - false_positives: 70.0000 - fals
e_negatives: 18.00 - 0s 2ms/step - loss: 0.3388 - accuracy: 0.8443 - true_positives:
5034.0000 - true_negatives: 1437.0000 - false_positives: 769.0000 - false_negatives:
424.0000 - val_loss: 0.6055 - val_accuracy: 0.7401 - val_true_positives: 1187.0000 -
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
73.0000
Epoch 78/100
13/13 [==============================] - ETA: 0s - loss: 0.3429 - accuracy: 0.8417 -
true_positives: 411.0000 - true_negatives: 94.0000 - false_positives: 71.0000 - false
_negatives: 24.000 - 0s 2ms/step - loss: 0.3334 - accuracy: 0.8460 - true_positives:
5022.0000 - true_negatives: 1462.0000 - false_positives: 744.0000 - false_negatives:
436.0000 - val_loss: 0.6161 - val_accuracy: 0.7390 - val_true_positives: 1221.0000 -
val_true_negatives: 195.0000 - val_false_positives: 361.0000 - val_false_negatives: 1
39.0000
Epoch 79/100
13/13 [==============================] - ETA: 0s - loss: 0.3066 - accuracy: 0.8700 -
true_positives: 436.0000 - true_negatives: 86.0000 - false_positives: 65.0000 - false
_negatives: 13.000 - 0s 2ms/step - loss: 0.3381 - accuracy: 0.8475 - true_positives:
5046.0000 - true_negatives: 1449.0000 - false_positives: 757.0000 - false_negatives:
412.0000 - val_loss: 0.6288 - val_accuracy: 0.7385 - val_true_positives: 1226.0000 -
val_true_negatives: 189.0000 - val_false_positives: 367.0000 - val_false_negatives: 1
34.0000
Epoch 80/100
13/13 [==============================] - ETA: 0s - loss: 0.2933 - accuracy: 0.8750 -
true_positives: 428.0000 - true_negatives: 97.0000 - false_positives: 57.0000 - false
_negatives: 18.000 - 0s 2ms/step - loss: 0.3389 - accuracy: 0.8436 - true_positives:
5059.0000 - true_negatives: 1406.0000 - false_positives: 800.0000 - false_negatives:
399.0000 - val_loss: 0.6156 - val_accuracy: 0.7432 - val_true_positives: 1169.0000 -
val_true_negatives: 255.0000 - val_false_positives: 301.0000 - val_false_negatives: 1
91.0000
Epoch 81/100
13/13 [==============================] - ETA: 0s - loss: 0.3291 - accuracy: 0.8417 -
true_positives: 402.0000 - true_negatives: 103.0000 - false_positives: 52.0000 - fals
e_negatives: 43.00 - 0s 2ms/step - loss: 0.3296 - accuracy: 0.8469 - true_positives:
5021.0000 - true_negatives: 1470.0000 - false_positives: 736.0000 - false_negatives:
437.0000 - val_loss: 0.6083 - val_accuracy: 0.7437 - val_true_positives: 1194.0000 -
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
66.0000
Epoch 82/100
13/13 [==============================] - ETA: 0s - loss: 0.2982 - accuracy: 0.8750 -
true_positives: 415.0000 - true_negatives: 110.0000 - false_positives: 51.0000 - fals
e_negatives: 24.00 - 0s 2ms/step - loss: 0.3295 - accuracy: 0.8497 - true_positives:
5050.0000 - true_negatives: 1462.0000 - false_positives: 744.0000 - false_negatives:
408.0000 - val_loss: 0.6171 - val_accuracy: 0.7401 - val_true_positives: 1177.0000 -
val_true_negatives: 241.0000 - val_false_positives: 315.0000 - val_false_negatives: 1
83.0000
Epoch 83/100
13/13 [==============================] - ETA: 0s - loss: 0.3068 - accuracy: 0.8650 -
true_positives: 403.0000 - true_negatives: 116.0000 - false_positives: 51.0000 - fals
e_negatives: 30.00 - 0s 2ms/step - loss: 0.3357 - accuracy: 0.8456 - true_positives:
```

```
5033.0000 - true_negatives: 1448.0000 - false_positives: 758.0000 - false_negatives: 
425.0000 - val_loss: 0.6276 - val_accuracy: 0.7406 - val_true_positives: 1137.0000 - 
val_true_negatives: 282.0000 - val_false_positives: 274.0000 - val_false_negatives: 2
23.0000
Epoch 84/100
13/13 [==============================] - ETA: 0s - loss: 0.3556 - accuracy: 0.8367 - 
true_positives: 379.0000 - true_negatives: 123.0000 - false_positives: 48.0000 - fals
e_negatives: 50.00 - 0s 1ms/step - loss: 0.3350 - accuracy: 0.8477 - true_positives: 
5048.0000 - true_negatives: 1449.0000 - false_positives: 757.0000 - false_negatives: 
410.0000 - val_loss: 0.6317 - val_accuracy: 0.7317 - val_true_positives: 1095.0000 - 
val_true_negatives: 307.0000 - val_false_positives: 249.0000 - val_false_negatives: 2
65.0000
Epoch 85/100
13/13 [==============================] - ETA: 0s - loss: 0.3563 - accuracy: 0.8350 - 
true_positives: 368.0000 - true_negatives: 133.0000 - false_positives: 39.0000 - fals
e_negatives: 60.00 - 0s 2ms/step - loss: 0.3350 - accuracy: 0.8454 - true_positives: 
5020.0000 - true_negatives: 1459.0000 - false_positives: 747.0000 - false_negatives: 
438.0000 - val_loss: 0.6300 - val_accuracy: 0.7317 - val_true_positives: 1101.0000 - 
val_true_negatives: 301.0000 - val_false_positives: 255.0000 - val_false_negatives: 2
59.0000
```

Out[67]:

```
<tensorflow.python.keras.callbacks.History at 0x1a1cb39ab48>
```

Create Dataframe for LIME

In [68]:

```python
df_train_lime = preprocess_df_for_lime(df_train_final1[features], numeric_features, bi
x_train_lime = df_train_lime.drop(LABEL, axis=1)
y_train_lime = df_train_lime[LABEL]

df_val_lime = preprocess_df_for_lime(df_val_final1[features], numeric_features, binary
x_val_lime = df_val_lime.drop(LABEL, axis=1)
y_val_lime = df_val_lime[LABEL]

df_test_lime = preprocess_df_for_lime(df_test_final1[features], numeric_features, bina
x_test_lime = df_test_lime.drop(LABEL, axis=1)
y_test_lime = df_test_lime[LABEL]
```

In [69]:

```python
print(x_train_lime.shape)
display(x_train_lime.head())
```

(7664, 9)

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_at
4310	0.815000	0.062178	1	9	0	3	0	
4311	0.709583	0.018187	0	6	0	0	0	
4312	0.795000	0.169789	1	8	2	4	0	
4313	0.748182	0.055315	0	8	1	4	0	
4314	0.602000	0.029579	1	8	0	2	0	

In [70]:

```python
cat_indices = [2, 3, 4, 5, 6, 8]

cat_names = {
```

```
    2:["F", "M"],
    3:['East Anglian Region', 'Scotland', 'South East Region',
     'West Midlands Region', 'Wales', 'North Western Region', 'South Region',
     'South West Region', 'East Midlands Region', 'Yorkshire Region',
     'London Region', 'North Region', 'Ireland'],
    4:['HE Qualification', 'A Level or Equivalent', 'Lower Than A Level',
     'Post Graduate Qualification', 'No Formal quals'],
    5:['80-100%', '60-80%', '40-60%', '20-40%', '0-20%'],
    6:['35-55', '0-35'],
    8:["Y", "N"]
}
```

In [71]:

```
explainer = lime.lime_tabular.LimeTabularExplainer(x_train_lime.values,
                                                   training_labels=y_train_lime.values
                                                   feature_names=x_train_lime.columns,
                                                   class_names=[FAIL, PASS],
                                                   mode="classification",
                                                   categorical_features=cat_indices,
                                                   categorical_names=cat_names,
                                                   kernel_width=3
                                                   )
```

In [72]:

```
test_convert = convert_lime_df_to_keras(x_test_lime, cat_values, binary)
test_predictions = model1.predict(test_convert.values)
```

In [73]:

```
import random
test_example = random.randint(0, len(x_test_lime))

row = x_test_lime.iloc[test_example]

display(df_test_final1[features].iloc[test_example])
print("Model prediction =", (FAIL if test_predictions[test_example] == 0 else PASS), t

def predict(instance):

    temp_df = pd.DataFrame(instance, columns = list(x_test_lime.columns))
    print("\nPerturbed features:")
    display(temp_df.head())
    print("\n")
    temp_df = convert_lime_df_to_keras(temp_df, cat_values, binary)

    predictions = model1.predict(temp_df.values)
    converted_predictions = []
    for p in predictions:
        two_vals = [1 - p[0], p[0]]
        converted_predictions.append(np.asarray(two_vals))
    return np.asarray(converted_predictions)

explanation = explainer.explain_instance(row, predict_fn=predict, num_features=len(x_t

explanation.show_in_notebook(show_table=True)
```

```
avg_score                          0.856047
sum_click                          0.049756
gender                                    F
region                 North Western Region
highest_education         Lower Than A Level
imd_band                             60-80%
age_band                              35-55
num_of_prev_attempts                      N
disability                                N
final_result                              1
Name: 1539, dtype: object
Model prediction = 1 [0.07938951]

Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.856047	0.049756	0.0	4.0	1.0	3.0	1.0	
1	0.267584	0.015246	0.0	4.0	0.0	3.0	1.0	
2	0.654891	0.296613	0.0	10.0	2.0	4.0	0.0	
3	0.891020	0.013429	0.0	5.0	1.0	3.0	0.0	
4	0.468602	0.007569	1.0	11.0	0.0	0.0	0.0	

Feature	Value
avg_score	0.86
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
gender=F	True
sum_click	0.05
region=Wales	True
disability=Y	True
imd_band=20-40%	True
age_band=0-35	True

In [74]:

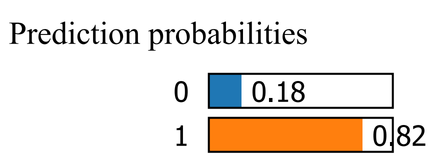
```
class_predictions = (test_predictions > 0.5).astype("int32")

false_positives = []
false_negatives = []

for i in range(len(x_test_lime)):
    if class_predictions[i][0] == 1 and y_test_lime.iloc[i] == 0:
        false_positives.append(x_test_lime.iloc[i])
    elif class_predictions[i][0] == 0 and y_test_lime.iloc[i] == 1:
        false_negatives.append(x_test_lime.iloc[i])

fp_examples = random.sample(range(0, len(false_positives)), 10)
fn_examples = random.sample(range(0, len(false_negatives)), 10)
```

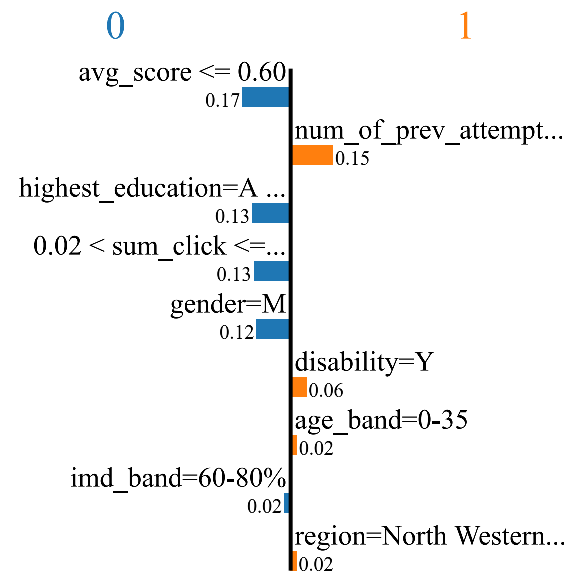
In [75]:

```
print("False Postives")
for fp in fp_examples:
    explanation = explainer.explain_instance(false_positives[fp],
                                             predict_fn=predict,
                                             num_features=len(x_test_lime.columns))
    explanation.show_in_notebook(show_table=True)
```

```
False Postives

Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.480000	0.022030	1.0	5.0	1.0	1.0	1.0	
1	0.643034	0.414014	0.0	5.0	0.0	3.0	1.0	
2	0.886464	0.006677	1.0	9.0	0.0	2.0	0.0	
3	0.865896	0.034874	0.0	5.0	1.0	0.0	1.0	
4	0.283975	0.023419	1.0	4.0	0.0	0.0	0.0	

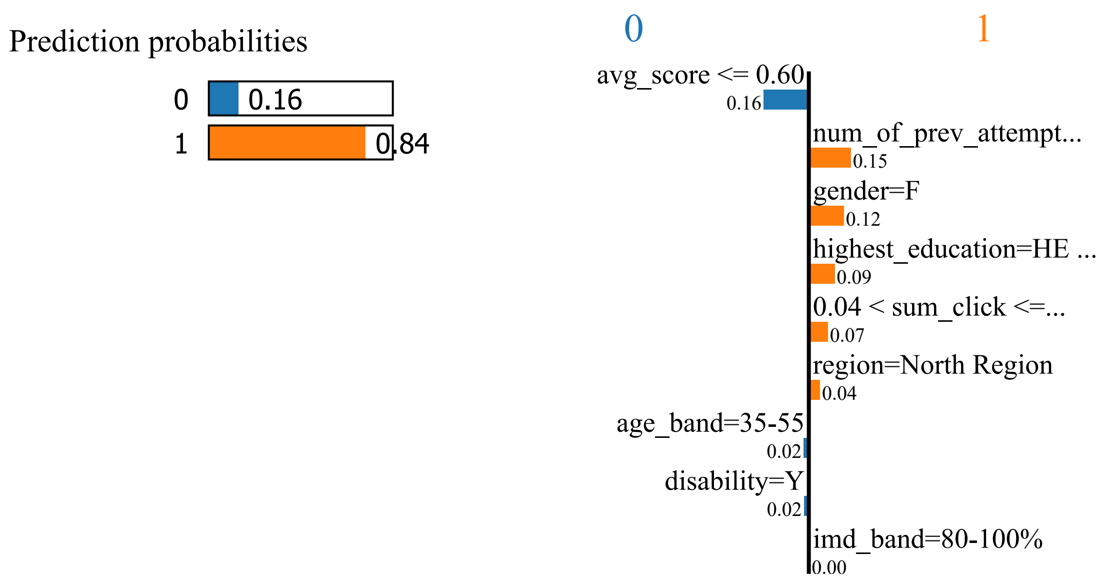




Feature	Value
avg_score	0.48
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
sum_click	0.02
gender=M	True
disability=Y	True
age_band=0-35	True
imd_band=60-80%	True
region=North Western Region	True

```
Perturbed features:
```

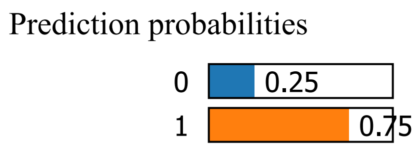
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.481667	0.079885	0.0	11.0	0.0	0.0	0.0	
1	0.798823	0.077202	0.0	10.0	1.0	3.0	0.0	
2	0.817223	0.005741	1.0	3.0	1.0	3.0	0.0	
3	0.641420	0.012020	1.0	4.0	2.0	0.0	1.0	
4	0.623247	0.063719	1.0	10.0	0.0	2.0	0.0	

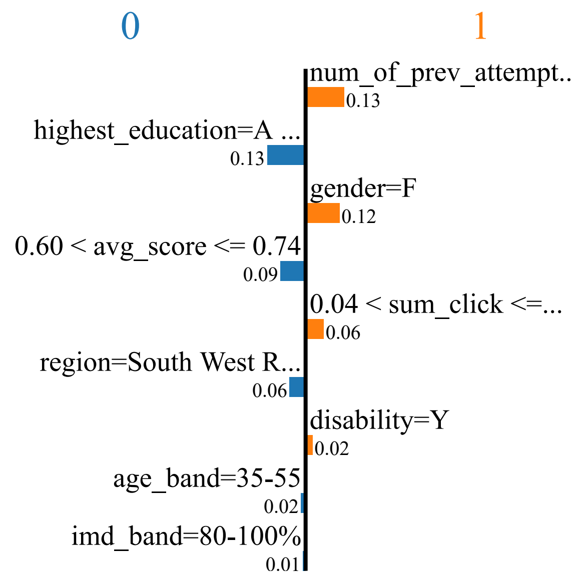


Feature	Value
avg_score	0.48
num_of_prev_attempts	0.00
gender=F	True
highest_education=HE Qualification	True
sum_click	0.08
region=North Region	True
age_band=35-55	True
disability=Y	True
imd_band=80-100%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.710909	0.047217	0.0	7.0	1.0	0.0	0.0	
1	0.893950	0.027598	0.0	10.0	4.0	4.0	0.0	
2	0.792395	0.164466	0.0	1.0	1.0	3.0	1.0	
3	0.870757	0.046214	1.0	2.0	0.0	1.0	0.0	
4	0.773448	0.072341	1.0	0.0	0.0	2.0	1.0	

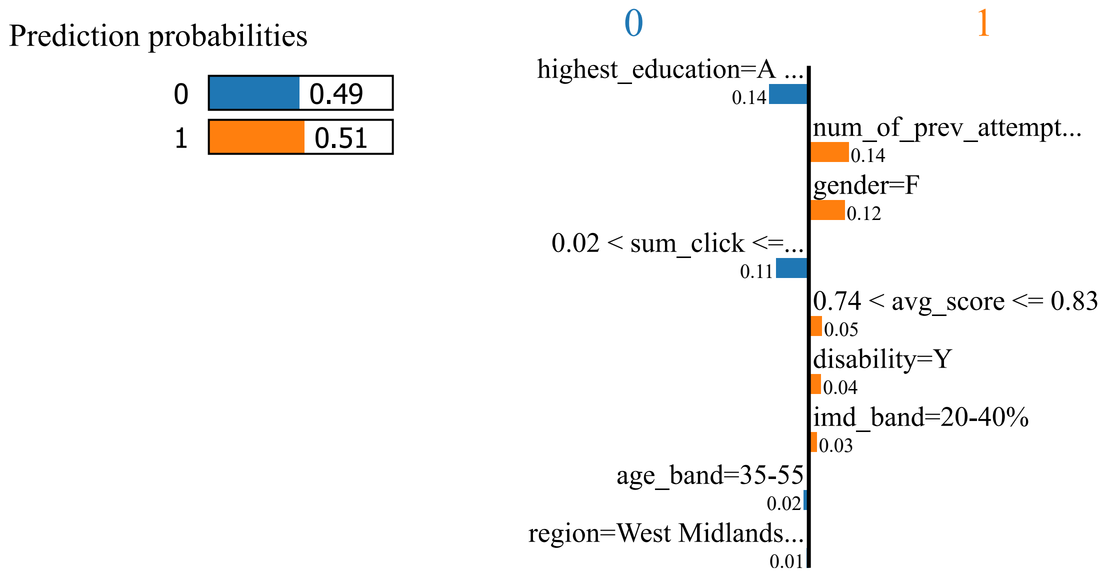




Feature	Value
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
gender=F	True
avg_score	0.71
sum_click	0.05
region=South West Region	True
disability=Y	True
age_band=35-55	True
imd_band=80-100%	True

```
Perturbed features:
```

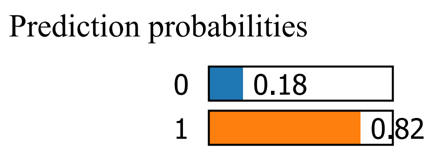
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.820000	0.035687	0.0	3.0	1.0	3.0	0.0	
1	0.183033	0.051423	1.0	6.0	0.0	1.0	0.0	
2	0.766229	0.116831	1.0	1.0	0.0	0.0	0.0	
3	0.668496	0.075587	0.0	10.0	0.0	3.0	0.0	
4	0.267872	0.000815	0.0	9.0	0.0	0.0	0.0	

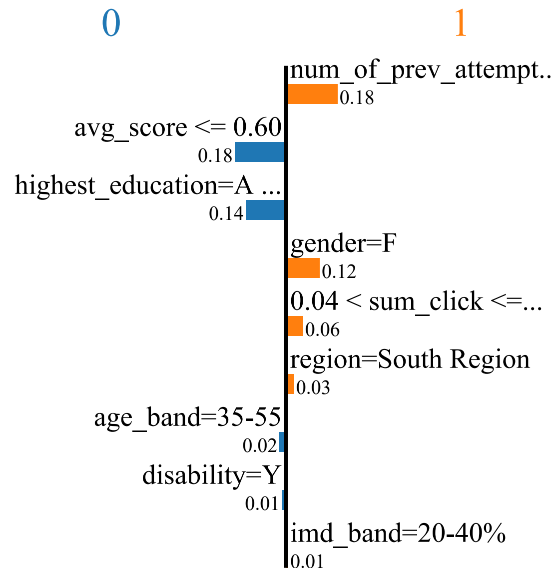


Feature	Value
highest_education=A Level or Equivalent	True
num_of_prev_attempts	0.00
gender=F	True
sum_click	0.04
avg_score	0.82
disability=Y	True
imd_band=20-40%	True
age_band=35-55	True
region=West Midlands Region	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.043374	0.0	6.0	1.0	3.0	0.0	
1	0.667461	0.237216	0.0	8.0	2.0	1.0	0.0	
2	0.105881	0.059210	1.0	0.0	0.0	2.0	0.0	
3	0.679205	0.010823	1.0	1.0	0.0	3.0	0.0	
4	0.274943	0.035627	0.0	2.0	0.0	3.0	0.0	

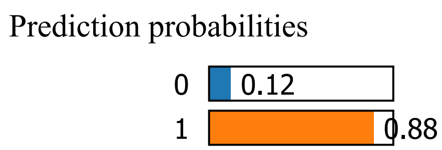


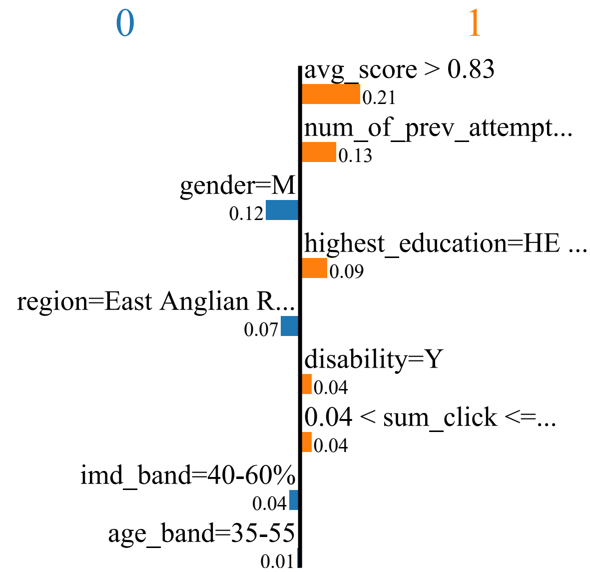


Feature	Value
num_of_prev_attempts	0.00
avg_score	0.00
highest_education=A Level or Equivalent	True
gender=F	True
sum_click	0.04
region=South Region	True
age_band=35-55	True
disability=Y	True
imd_band=20-40%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.944545	0.074394	1.0	0.0	0.0	2.0	0.0	
1	0.694932	0.005782	1.0	7.0	0.0	2.0	1.0	
2	0.886838	0.027059	1.0	6.0	0.0	4.0	0.0	
3	0.039645	0.064627	1.0	4.0	0.0	1.0	1.0	
4	0.897133	0.031089	0.0	2.0	2.0	1.0	0.0	

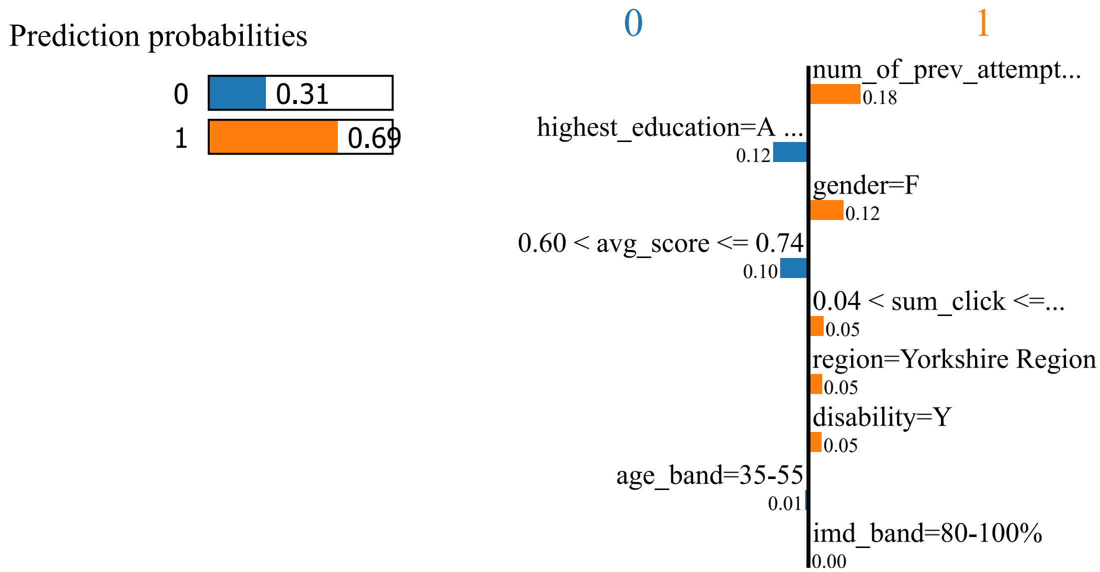




Feature	Value
avg_score	0.94
num_of_prev_attempts	0.00
gender=M	True
highest_education=HE Qualification	True
region=East Anglian Region	True
disability=Y	True
sum_click	0.07
imd_band=40-60%	True
age_band=35-55	True

```
Perturbed features:
```

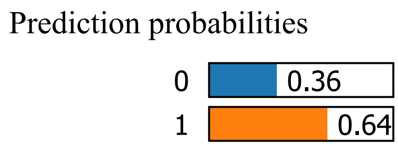
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.656667	0.043168	0.0	9.0	1.0	0.0	0.0	
1	0.847873	0.016666	0.0	4.0	2.0	0.0	0.0	
2	0.365556	0.096685	1.0	3.0	0.0	2.0	0.0	
3	0.778315	0.032380	0.0	3.0	0.0	3.0	0.0	
4	0.348489	0.026593	1.0	5.0	2.0	1.0	1.0	

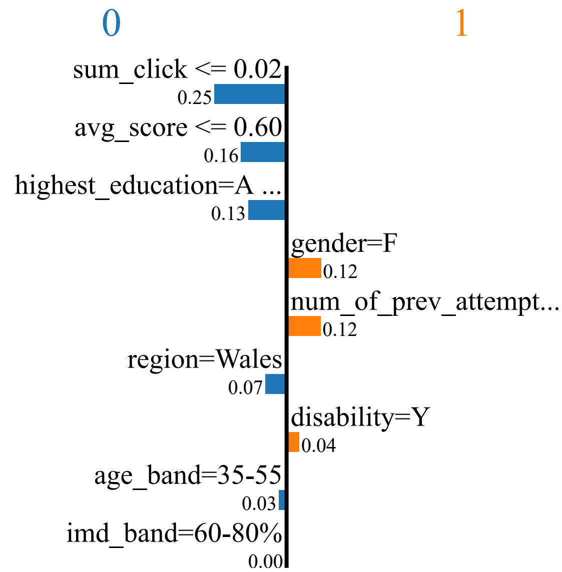


Feature	Value
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
gender=F	True
avg_score	0.66
sum_click	0.04
region=Yorkshire Region	True
disability=Y	True
age_band=35-55	True
imd_band=80-100%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.011667	0.0	4.0	1.0	1.0	0.0	
1	0.062739	0.000665	1.0	6.0	2.0	3.0	0.0	
2	0.658782	0.011194	0.0	5.0	0.0	4.0	0.0	
3	0.699997	0.040609	1.0	8.0	1.0	0.0	0.0	
4	0.685703	0.023340	1.0	1.0	0.0	1.0	0.0	

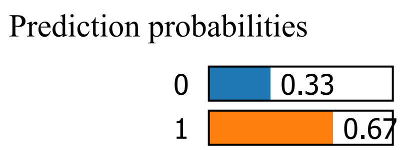


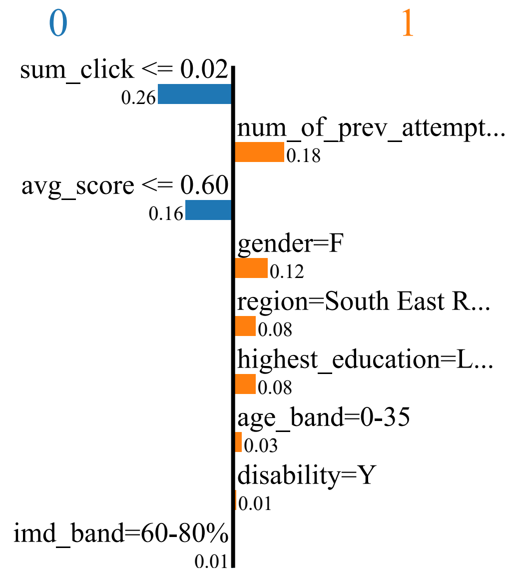


Feature	Value
sum_click	0.01
avg_score	0.00
highest_education=A Level or Equivalent	True
gender=F	True
num_of_prev_attempts	0.00
region=Wales	True
disability=Y	True
age_band=35-55	True
imd_band=60-80%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.007618	0.0	2.0	2.0	1.0	1.0	
1	0.002691	0.156219	0.0	1.0	3.0	0.0	0.0	
2	0.930816	0.027679	0.0	12.0	1.0	1.0	1.0	
3	0.845543	0.010512	1.0	2.0	0.0	3.0	0.0	
4	0.552066	0.054104	1.0	8.0	0.0	1.0	1.0	

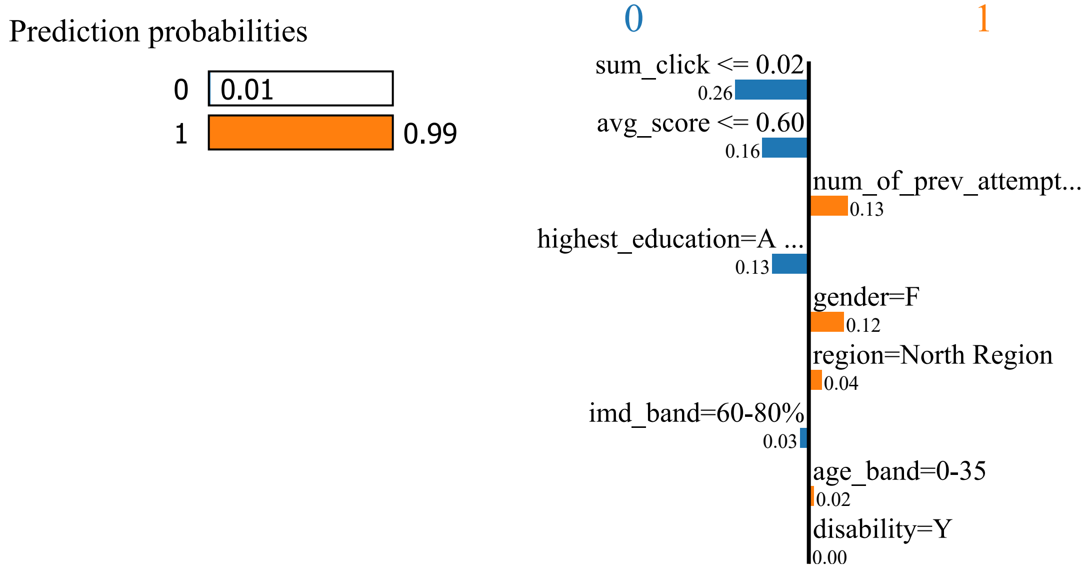




Feature	Value
sum_click	0.01
num_of_prev_attempts	0.00
avg_score	0.00
gender=F	True
region=South East Region	True
highest_education=Lower Than A Level	True
age_band=0-35	True
disability=Y	True
imd_band=60-80%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.000549	0.0	11.0	1.0	1.0	1.0	
1	0.197236	0.026246	1.0	1.0	0.0	3.0	1.0	
2	0.781088	0.195866	0.0	6.0	1.0	2.0	0.0	
3	0.641892	0.016094	1.0	3.0	2.0	3.0	0.0	
4	0.895385	0.063972	0.0	0.0	1.0	3.0	0.0	



Feature	Value
sum_click	0.00
avg_score	0.00
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
gender=F	True
region=North Region	True
imd_band=60-80%	True
age_band=0-35	True
disability=Y	True

In the first case: there are three features that might be correlated with the false positive predictions is sum_click, num_of_prev_attempt and disability

in the second case: highest education, number of prev attemp, region, disability and avg_score are the factors correlated to false possitive prediction

in the third case, highest_edu, num_of_prev_attempts and disability are correlated with the false positive predictions

in the 4th case, sum_click and disability are correlated with the false positive result

in the 4th case, num_of_pre_attemps,sum_click,disability and imd_band are correlated with the false positive result

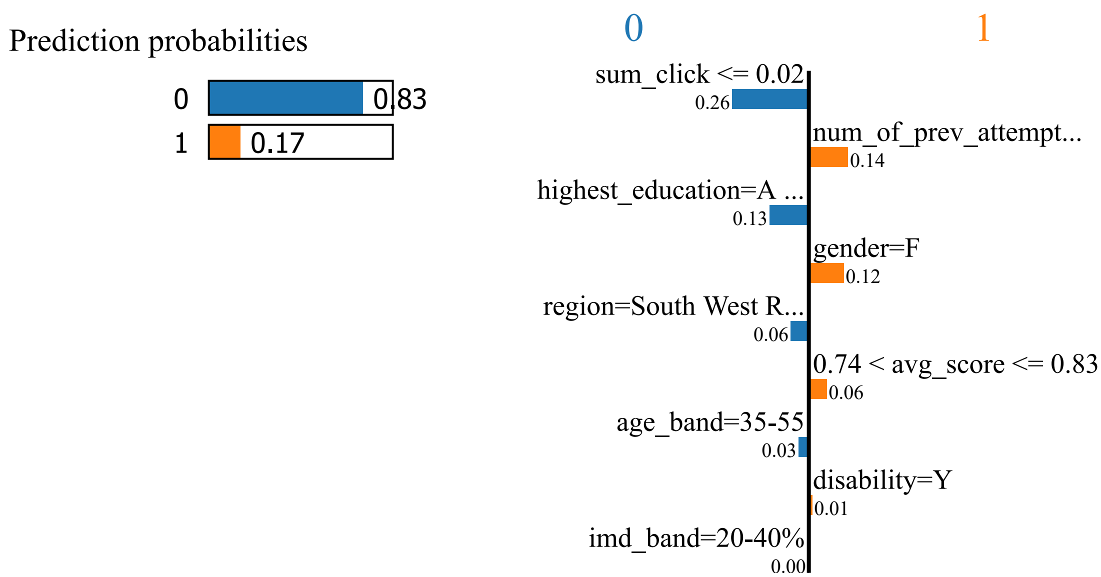
Inconclusion, predictions based on num_of_prev_attempts and sum_click seemingly result in false positive predictions

In [76]:
```
print("False Negatives")
```

```
for fn in fn_examples:
    explanation = explainer.explain_instance(false_negatives[fn],
                                             predict_fn=predict,
                                             num_features=len(x_test_lime.columns))
    explanation.show_in_notebook(show_table=True)
```

```
False Negatives

Perturbed features:
```

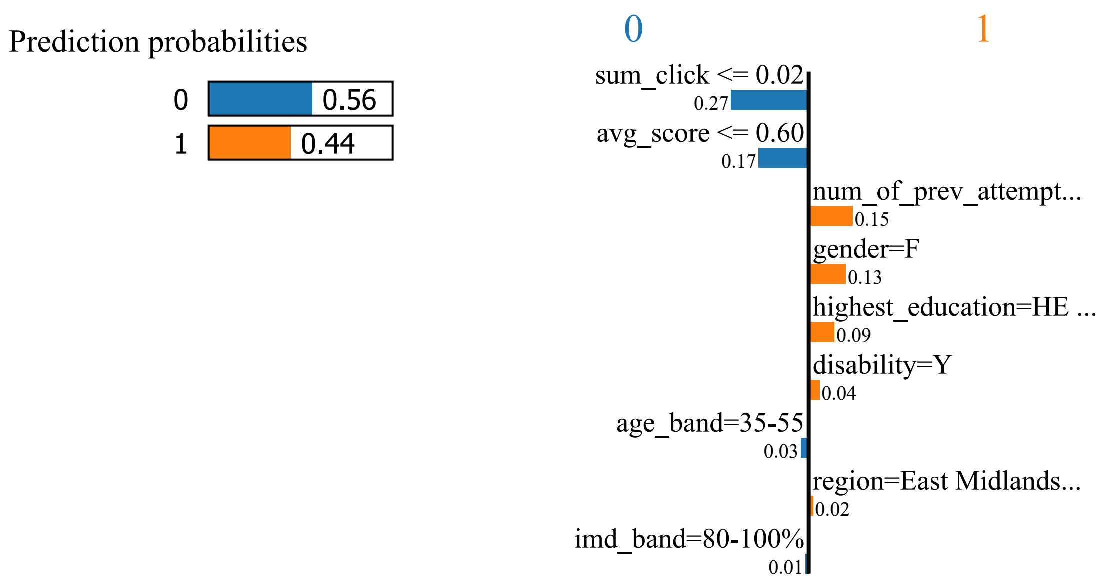
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.740000	0.005079	0.0	7.0	1.0	3.0	0.0	
1	0.889958	0.017986	0.0	2.0	2.0	4.0	0.0	
2	0.520147	0.208771	1.0	7.0	0.0	4.0	1.0	
3	0.628418	0.265428	1.0	0.0	2.0	2.0	0.0	
4	0.743510	0.049656	0.0	1.0	2.0	2.0	0.0	



Feature	Value
sum_click	0.01
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
gender=F	True
region=South West Region	True
avg_score	0.74
age_band=35-55	True
disability=Y	True
imd_band=20-40%	True

```
Perturbed features:
```

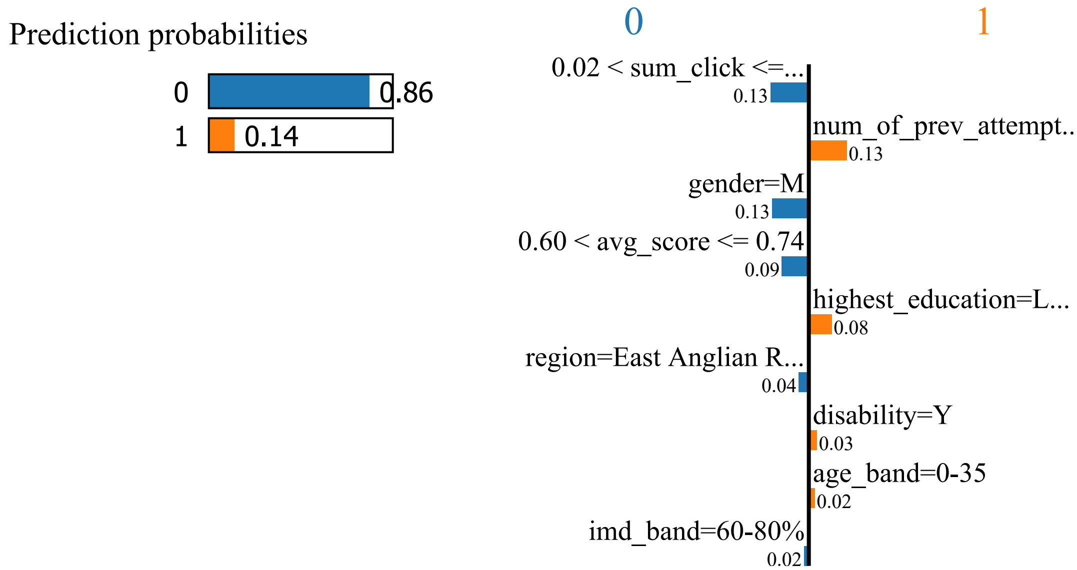
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.008098	0.0	8.0	0.0	0.0	0.0	
1	0.881936	0.032467	1.0	1.0	0.0	1.0	0.0	
2	0.621713	0.034972	0.0	10.0	1.0	0.0	0.0	
3	0.677342	0.048690	1.0	3.0	1.0	4.0	1.0	
4	0.785962	0.025356	0.0	9.0	0.0	4.0	0.0	



Feature	Value
sum_click	0.01
avg_score	0.00
num_of_prev_attempts	0.00
gender=F	True
highest_education=HE Qualification	True
disability=Y	True
age_band=35-55	True
region=East Midlands Region	True
imd_band=80-100%	True

```
Perturbed features:
```

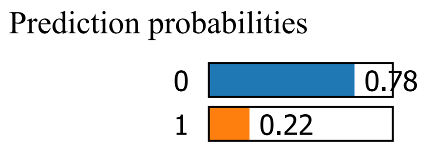
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.616667	0.040903	1.0	0.0	2.0	1.0	1.0	
1	0.099940	0.001017	0.0	0.0	2.0	2.0	0.0	
2	0.214092	0.036261	0.0	11.0	1.0	2.0	0.0	
3	0.686724	0.048262	0.0	0.0	1.0	3.0	0.0	
4	0.776763	0.030608	0.0	0.0	1.0	0.0	0.0	

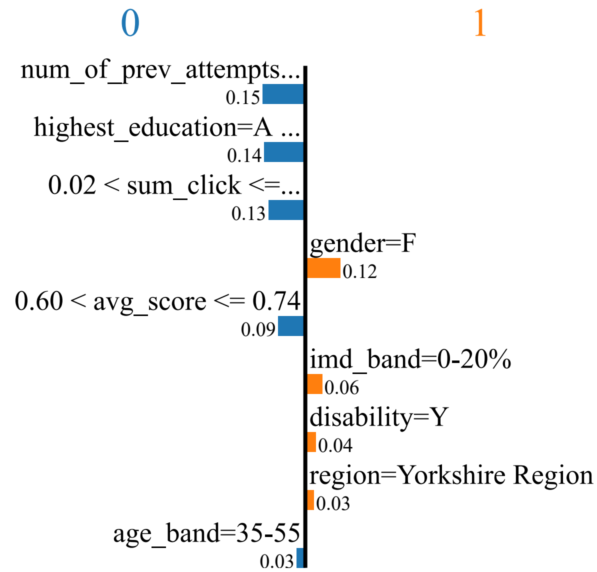


Feature	Value
sum_click	0.04
num_of_prev_attempts	0.00
gender=M	True
avg_score	0.62
highest_education=Lower Than A Level	True
region=East Anglian Region	True
disability=Y	True
age_band=0-35	True
imd_band=60-80%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.729167	0.027932	0.0	9.0	1.0	4.0	0.0	
1	0.693459	0.039723	1.0	3.0	0.0	2.0	0.0	
2	0.871326	0.124373	0.0	8.0	1.0	3.0	0.0	
3	0.686231	0.067025	1.0	0.0	1.0	4.0	0.0	
4	0.781209	0.029230	0.0	7.0	0.0	1.0	0.0	

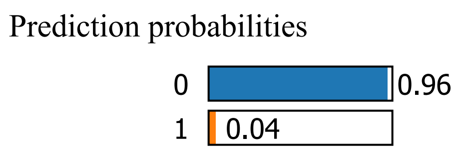


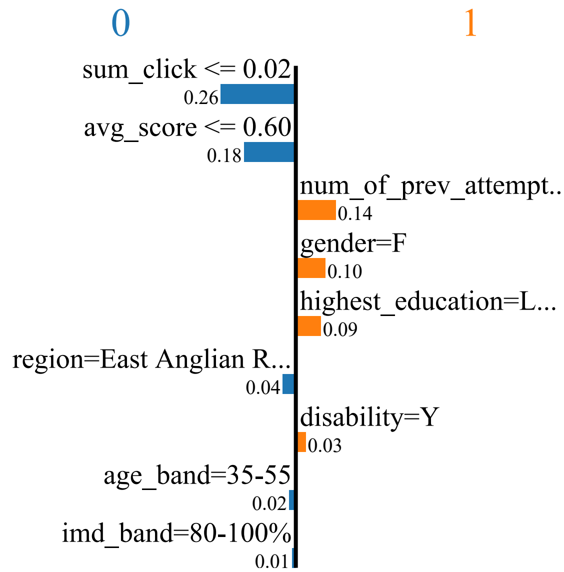


Feature	Value
num_of_prev_attempts	1.00
highest_education=A Level or Equivalent	True
sum_click	0.03
gender=F	True
avg_score	0.73
imd_band=0-20%	True
disability=Y	True
region=Yorkshire Region	True
age_band=35-55	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.016471	0.0	0.0	2.0	0.0	0.0	
1	0.903358	0.006864	1.0	2.0	2.0	1.0	1.0	
2	0.764698	0.006621	1.0	8.0	2.0	2.0	0.0	
3	0.791392	0.210003	0.0	9.0	1.0	1.0	0.0	
4	0.862237	0.034020	0.0	2.0	0.0	1.0	0.0	

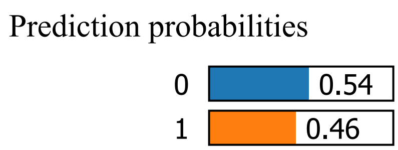


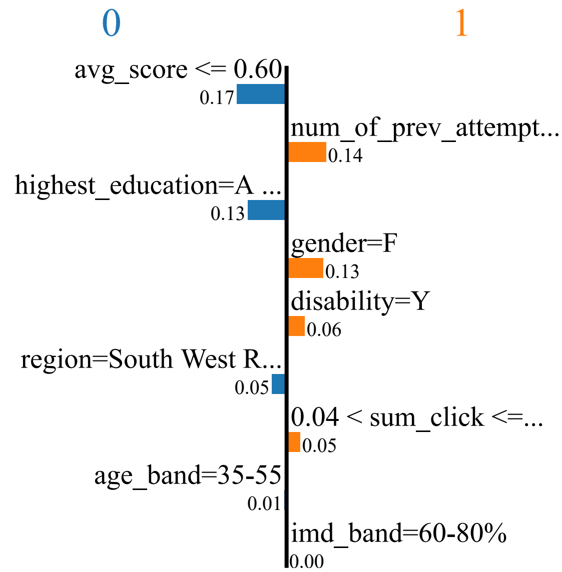


Feature	Value
sum_click	0.02
avg_score	0.00
num_of_prev_attempts	0.00
gender=F	True
highest_education=Lower Than A Level	True
region=East Anglian Region	True
disability=Y	True
age_band=35-55	True
imd_band=80-100%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.536667	0.043717	0.0	7.0	1.0	1.0	0.0	
1	0.869927	0.268735	1.0	7.0	0.0	0.0	0.0	
2	0.668643	0.039658	0.0	3.0	0.0	3.0	0.0	
3	0.468894	0.024804	1.0	7.0	2.0	1.0	0.0	
4	0.910850	0.104262	1.0	11.0	0.0	3.0	0.0	

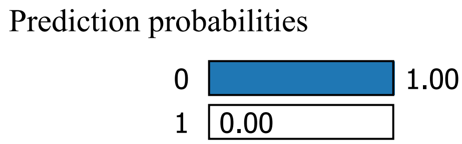


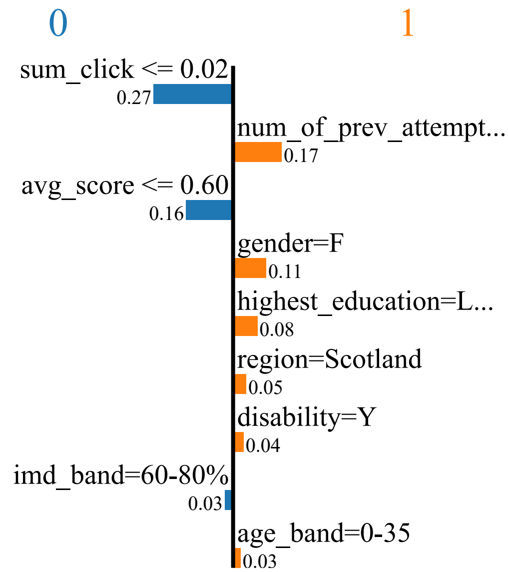


Feature	Value
avg_score	0.54
num_of_prev_attempts	0.00
highest_education=A Level or Equivalent	True
gender=F	True
disability=Y	True
region=South West Region	True
sum_click	0.04
age_band=35-55	True
imd_band=60-80%	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.000000	0.008167	0.0	1.0	2.0	1.0	1.0	
1	0.886387	0.060652	0.0	11.0	2.0	1.0	0.0	
2	0.441151	0.014870	0.0	10.0	1.0	3.0	1.0	
3	0.893444	0.008881	0.0	2.0	1.0	2.0	0.0	
4	0.683864	0.007907	0.0	0.0	1.0	1.0	1.0	

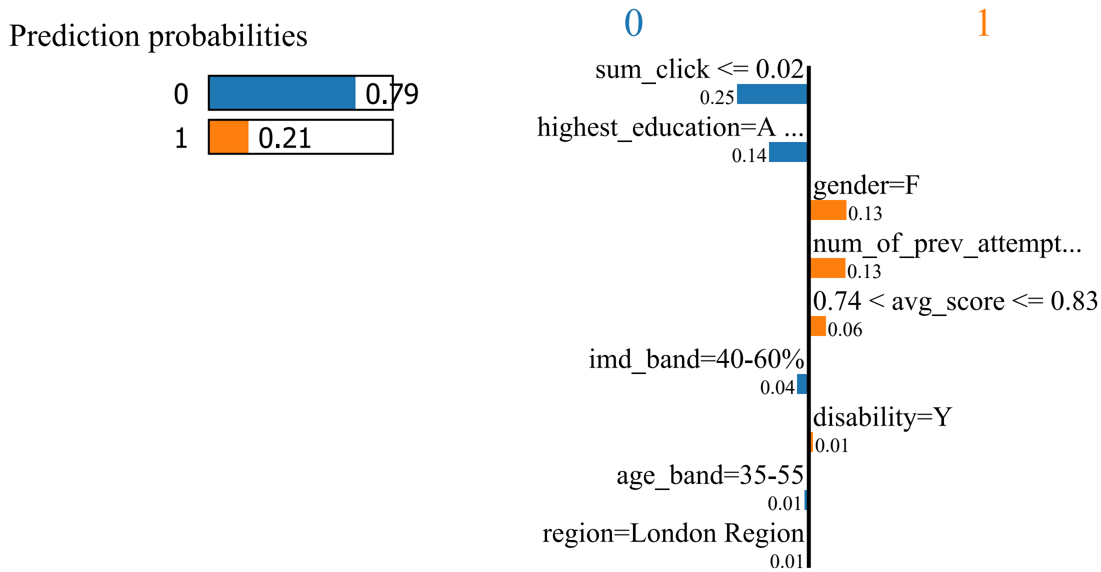




Feature	Value
sum_click	0.01
num_of_prev_attempts	0.00
avg_score	0.00
gender=F	True
highest_education=Lower Than A Level	True
region=Scotland	True
disability=Y	True
imd_band=60-80%	True
age_band=0-35	True

```
Perturbed features:
```

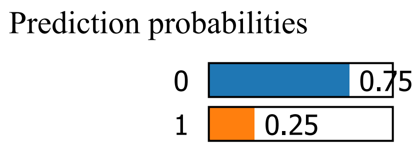
	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.736000	0.010638	0.0	10.0	1.0	2.0	0.0	
1	0.843021	0.063456	1.0	2.0	1.0	2.0	0.0	
2	0.811959	0.023075	1.0	8.0	2.0	3.0	1.0	
3	0.702888	0.045481	0.0	5.0	0.0	4.0	0.0	
4	0.780927	0.034264	0.0	12.0	1.0	0.0	0.0	

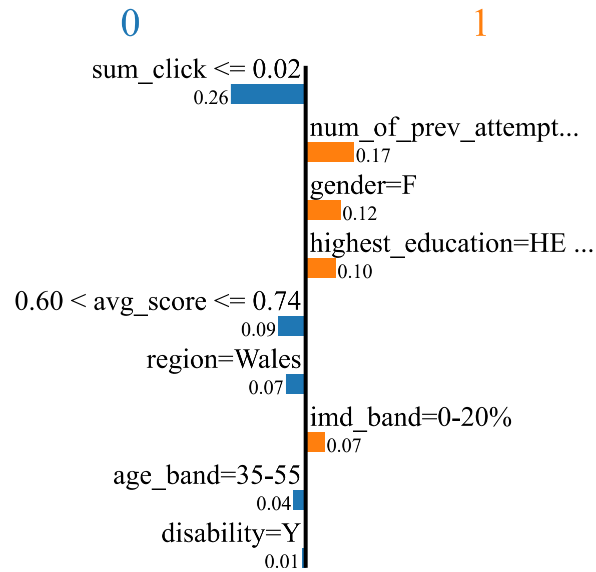


Feature	Value
sum_click	0.01
highest_education=A Level or Equivalent	True
gender=F	True
num_of_prev_attempts	0.00
avg_score	0.74
imd_band=40-60%	True
disability=Y	True
age_band=35-55	True
region=London Region	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.721000	0.012147	0.0	4.0	0.0	4.0	0.0	
1	0.886542	0.038767	1.0	2.0	0.0	0.0	0.0	
2	0.562807	0.164860	1.0	6.0	1.0	3.0	0.0	
3	0.336532	0.010262	1.0	6.0	0.0	2.0	0.0	
4	0.885100	0.161311	0.0	1.0	1.0	1.0	0.0	

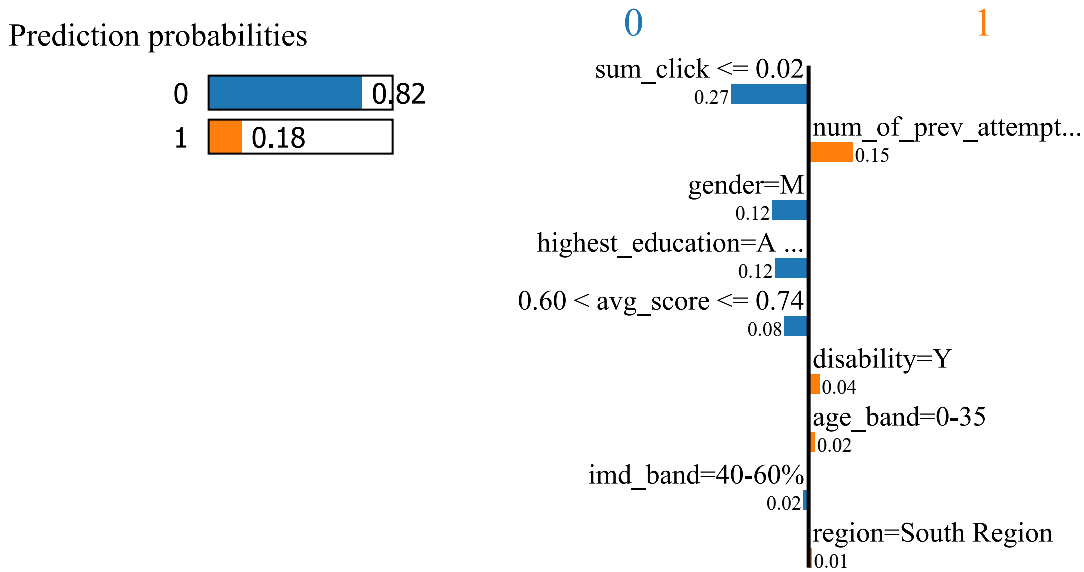




Feature	Value
sum_click	0.01
num_of_prev_attempts	0.00
gender=F	True
highest_education=HE Qualification	True
avg_score	0.72
region=Wales	True
imd_band=0-20%	True
age_band=35-55	True
disability=Y	True

```
Perturbed features:
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.647143	0.005010	1.0	6.0	1.0	2.0	1.0	
1	0.706384	0.142513	1.0	8.0	1.0	0.0	0.0	
2	0.447063	0.005094	0.0	5.0	0.0	1.0	0.0	
3	0.688829	0.040763	0.0	0.0	2.0	3.0	0.0	
4	0.898423	0.008161	1.0	6.0	0.0	2.0	0.0	



Feature	Value
sum_click	0.01
num_of_prev_attempts	0.00
gender=M	True
highest_education=A Level or Equivalent	True
avg_score	0.65
disability=Y	True
age_band=0-35	True
imd_band=40-60%	True
region=South Region	True

case 1: highest education, gender, avg_score and region are features correlated to false negative case 2: highest education, gender, avg_score and imd_band are features correlated to false negative case 3: sum_click, gender, avg_score and gender are features correlated to false negative case 4: sum_click, gender, avg_score and gender are the most likely features correlated to false negative case 5: sum_click avg_score are most likely features correlated to false negative

From those examples, I think gender and avg_score are the most common features resulting in false negative result

In [77]:
```
df_test_final1.head()
```

Out[77]:

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_atte
0	0.616667	0.014069	M	South West Region	A Level or Equivalent	20-40%	0-35	
1	0.885000	0.143230	M	Yorkshire Region	No Formal quals	0-20%	0-35	
2	0.805000	0.129916	M	North Western Region	Lower Than A Level	40-60%	0-35	
3	0.821250	0.145769	F	West Midlands Region	A Level or Equivalent	20-40%	0-35	
4	0.850000	0.011598	M	East Midlands Region	A Level or Equivalent	80-100%	0-35	

In [78]:

```
def label_encode_df(df, categories):
    '''
      Label encodes all categorical features (including binary).
      - df: The dataframe to encode
      - categories: a dictionary. Keys correspond to df column names. Values are
        lists of column values.
    '''
    new_df = df.copy(deep=True)
    for feature in df.columns:
        if feature in categories.keys():
            new_df[feature] = df[feature].apply(lambda x: cat_values[feature].index(x)
    return new_df
```

In [79]:

```
df_test_label_encoded = label_encode_df(df_test_final1[features], cat_values)

display(df_test_label_encoded.head())
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_attem
0	0.616667	0.014069	1	5	0	0	0	
1	0.885000	0.143230	1	10	3	2	0	
2	0.805000	0.129916	1	4	1	1	0	
3	0.821250	0.145769	0	6	0	0	0	
4	0.850000	0.011598	1	8	0	4	0	

In [80]:

```
dtree = DecisionTreeClassifier()
x_test_label_encoded = df_test_label_encoded.drop(LABEL, axis=1)
dtree.fit(x_test_label_encoded, (test_predictions > 0.5).astype("int32"))
```

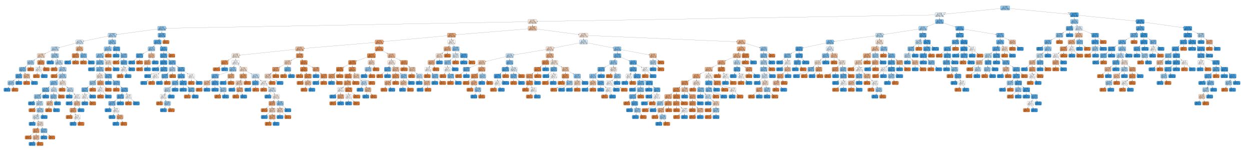
Out[80]: DecisionTreeClassifier()

In [81]:

```
dot_data = StringIO()
```

```
export_graphviz(dtree, out_file=dot_data, filled=True, rounded=True, special_character
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
Image(graph.create_png())
```

Out[81]:



In [82]:

```
print("all features")
for i, feature in enumerate(x_test_label_encoded.columns):
    print(i,":",feature)
print("\n")
print("category features")
for feature in cat_values.keys():
    print(feature, ":", cat_values[feature])
```

```
all features
0 : avg_score
1 : sum_click
2 : gender
3 : region
4 : highest_education
5 : imd_band
6 : age_band
7 : num_of_prev_attempts
8 : disability


category features
region : ['Scotland', 'East Anglian Region', 'South Region', 'London Region', 'North
Western Region', 'South West Region', 'West Midlands Region', 'Wales', 'East Midlands
Region', 'South East Region', 'Yorkshire Region', 'Ireland', 'North Region']
highest_education : ['A Level or Equivalent', 'Lower Than A Level', 'HE Qualificatio
n', 'No Formal quals', 'Post Graduate Qualification']
imd_band : ['20-40%', '40-60%', '0-20%', '60-80%', '80-100%']
age_band : ['0-35', '35-55']
num_of_prev_attempts : ['N', 'Y']
gender : ['F', 'M']
disability : ['Y', 'N']
```

In [83]:

```
random_rows = random.sample(range(0, len(df_test_label_encoded)), 5)
df_random_with_preds = df_test_label_encoded[features].iloc[random_rows]
df_random_with_preds["predictions"] = (test_predictions > 0.5).astype("int32")[random_
display(df_random_with_preds)
```

	avg_score	sum_click	gender	region	highest_education	imd_band	age_band	num_of_prev_at
2391	0.000000	0.000755	1	4	2	2	0	
404	0.000000	0.083797	0	0	2	4	0	
1081	0.735698	0.007000	0	9	1	1	1	
673	0.823953	0.005765	1	4	1	4	1	
2379	0.865000	0.014275	1	7	0	2	0	

Explanation for case 1 and 2:

Case 1: Because the sum_click (0.052) is not less or equal to 0.045 it goes to avg_score Because the avg_score (0.79) is not less or equal to 0.635 it goes to avg_score Because the avg_score (0.79) is not less or equal to 0.745 it goes to highest_education Because highest_education(1) is less or equal to 2.5 it goes to sum_click because sum-click(0.052) is smaller than 0.06 it goes to disability because disability(1) is not less or equal to 0.5 it goes to region because region (0) is less than 11 it goes to sum_click because sum_click(0.052) is smaller than 0.057 it goes to age because age(1) is not smaller than 0.5 it goes to gender because gender (0) is not less than 0.5, it reaches the predictions.

case 2: Because the sum_click(0.027) is less than 0.045 it goes to avg_score Because the avg_score (0.69) is less or equal to 0.736 it goes to avg_score Because the avg_score (0.69) is less or equal to 0.736 it goes to imd_band Because imd_band (1) is less or equal to 2.5 it goes to gender because gender (0) is smaller than 0.5 it goes to imd_band Because imd_band (1) is less or equal to 1.5 it goes to region because region (7) is less than 0.5 it goes to num_of_attempts because num_of_attempts (0) is less than 0.5 it goes to disability because disability(1) is not less than 0.5 it goes to age because age(1) is not less than 0.5 it goes to sum_click because sum_click(0.027) is equal to 0.027 it goes to region because region (7) is less than 8.5 it goes to highest education Because highest_education(1) is less or equal to 1.5 it goes to sum_click because sum_click(0.027) is not less or equal to 0.02 it goes to sum_click because sum_click(0.027) is not less or equal to 0.022 it reaches the predictions

DICE

In [84]:
```
!pip install dice-ml
import dice_ml
```

```
Requirement already satisfied: dice-ml in c:\users\hongt\anaconda3\lib\site-packages
(0.6.1)
Requirement already satisfied: h5py in c:\users\hongt\anaconda3\lib\site-packages (fr
om dice-ml) (2.10.0)
Requirement already satisfied: numpy in c:\users\hongt\anaconda3\lib\site-packages (f
rom dice-ml) (1.20.1)
Requirement already satisfied: tqdm in c:\users\hongt\anaconda3\lib\site-packages (fr
om dice-ml) (4.59.0)
Requirement already satisfied: pandas in c:\users\hongt\anaconda3\lib\site-packages
(from dice-ml) (1.2.4)
Requirement already satisfied: scikit-learn in c:\users\hongt\anaconda3\lib\site-pack
ages (from dice-ml) (0.24.1)
Requirement already satisfied: six in c:\users\hongt\anaconda3\lib\site-packages (fro
m h5py->dice-ml) (1.15.0)
Requirement already satisfied: pytz>=2017.3 in c:\users\hongt\anaconda3\lib\site-pack
ages (from pandas->dice-ml) (2021.1)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\hongt\anaconda3\lib
\site-packages (from pandas->dice-ml) (2.8.1)
Requirement already satisfied: scipy>=0.19.1 in c:\users\hongt\anaconda3\lib\site-pac
kages (from scikit-learn->dice-ml) (1.6.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\hongt\anaconda3\lib\s
ite-packages (from scikit-learn->dice-ml) (2.1.0)
Requirement already satisfied: joblib>=0.11 in c:\users\hongt\anaconda3\lib\site-pack
ages (from scikit-learn->dice-ml) (1.0.1)
```

In [85]:

```
def convert_num_to_cat(x):
    if x <= 800:
        return "0-800"
    elif 800 < x and x <= 1600:
        return "801-1600"
    elif 1600 < x and x <= 2400:
        return "1601-2400"
    elif 2400 < x and x <= 3200:
        return "2401-3200"
    else:
        return ">3200"

org_final_df["sum_click"] = org_final_df["sum_click"].apply(lambda x: convert_num_to_c

display(org_final_df.head())
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	stu
1	69.2	801-1600	F	HE Qualification	20-40%	35-55	N	
2	72.4	801-1600	F	A Level or Equivalent	40-60%	35-55	N	
3	51.0	0-800	F	Lower Than A Level	40-60%	0-35	N	
4	73.0	801-1600	M	A Level or Equivalent	80-100%	35-55	N	
5	64.8	801-1600	M	HE Qualification	20-40%	0-35	N	

In [86]:

```
org_final_df= org_final_df.drop(columns = 'studied_credits')
new_org_final_df = org_final_df.sample(frac=1, random_state=32).reset_index(drop=True)

#Create training df and testing df
```

```
test_size_org_df = int(len(new_org_final_df) * 0.2)
df_test_org = new_org_final_df[:test_size_org_df]
df_train_val = new_org_final_df[test_size_org_df:]

#create training df and validating df
val_size_org = int(len(df_train_val) * 0.2)
df_val_org = df_train_val[:val_size_org]
df_train_org = df_train_val[val_size_org:]

#check if they are created as expected
print(df_train_org.shape)
display(df_train_org.head())
print(df_val_org.shape)
display(df_val_org.head())
print(df_test_org.shape)
display(df_test_org.head())
```

(7664, 9)

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts
4310	81.500000	801-1600	M	A Level or Equivalent	60-80%	0-35	N
4311	70.958333	0-800	F	A Level or Equivalent	20-40%	0-35	N
4312	79.500000	2401-3200	M	HE Qualification	80-100%	0-35	N
4313	74.818182	801-1600	F	Lower Than A Level	80-100%	0-35	N
4314	60.200000	0-800	M	A Level or Equivalent	0-20%	0-35	N

(1916, 9)

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts
2394	82.883721	0-800	F	A Level or Equivalent	20-40%	35-55	N
2395	52.000000	2401-3200	M	A Level or Equivalent	0-20%	0-35	N
2396	73.500000	2401-3200	M	A Level or Equivalent	60-80%	35-55	Y
2397	84.500000	801-1600	M	A Level or Equivalent	40-60%	0-35	N
2398	72.500000	1601-2400	M	HE Qualification	20-40%	0-35	N

(2394, 9)

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	61.666667	0-800	M	A Level or Equivalent	20-40%	0-35	N	
1	88.500000	1601-2400	M	No Formal quals	0-20%	0-35	N	
2	80.500000	1601-2400	M	Lower Than A Level	40-60%	0-35	N	
3	82.125000	1601-2400	F	A Level or Equivalent	20-40%	0-35	Y	
4	85.000000	0-800	M	A Level or Equivalent	80-100%	0-35	N	

In [87]:
```
# These are the features kept for this notebook
org_features = df_train_org.columns.tolist()

org_num_features = ["avg_score"]

# categorical features will be encoded seperatedly
org_cat_features = ["sum_click", "gender",  "highest_education", "imd_band", "age_banc
```

In [88]:
```
print(org_features)
```

```
['avg_score', 'sum_click', 'gender', 'highest_education', 'imd_band', 'age_band', 'nu
m_of_prev_attempts', 'disability', 'final_result']
```

In [89]:
```
#
num_values1 = {}
num_values1["avg_score"] = { "min": df_train_org["avg_score"].min(),
                             "max": df_train_org["avg_score"].max()
                           }
print(num_values1)

#create a dictionary to keep track the unique values of each columns.
cat_values1 = {}

for feature in org_cat_features:
    cat_values1[feature] = list(df_train_org[feature].value_counts().index)

cat_values1["gender"] = ["F", "M"]
cat_values1["disability"] = ["Y", "N"]
cat_values1["num_of_prev_attempts"] = ["Y", "N"]

print("\n")
print("category features")
for k,v in cat_values1.items():
    print(k,v)
```

```
{'avg_score': {'min': 0.0, 'max': 100.0}}


category features
sum_click ['0-800', '801-1600', '1601-2400', '2401-3200', '>3200']
gender ['F', 'M']
highest_education ['A Level or Equivalent', 'Lower Than A Level', 'HE Qualification',
'No Formal quals', 'Post Graduate Qualification']
imd_band ['20-40%', '40-60%', '0-20%', '60-80%', '80-100%']
age_band ['0-35', '35-55']
num_of_prev_attempts ['Y', 'N']
disability ['Y', 'N']
```

In [90]:

```
def preprocess_df_label(org_fin_df, numeric_features, num_values, cat_values):
    #new_df = org_fin_df.copy(deep=True)
    new_df = pd.get_dummies(org_fin_df, drop_first = False, columns = cat_values )
    for col in org_fin_df.columns:
        if col == 'avg_score':
            #new_df[col] = pd.to_numeric(org_fin_df[col])
            new_df[col] = new_df[col].apply(lambda x: rescale(x, new_df[col].min(), ne
        elif col == "final_result":
            new_df[col] = pd.get_dummies(org_fin_df[col], drop_first=True)
    return new_df
```

In [91]:

```
preprocess_train_df = preprocess_df_label(df_train_org, org_num_features, num_values1,
preprocess_val_df = preprocess_df_label(df_val_org, org_num_features,num_values1, cat_
preprocess_test_df = preprocess_df_label(df_test_org, org_num_features, num_values1, c

NO_LABEL = preprocess_train_df.columns.tolist()
NO_LABEL.remove(LABEL)

x_train_dice = preprocess_train_df[NO_LABEL]
y_train_dice = preprocess_train_df[LABEL]

x_val_dice = preprocess_val_df[NO_LABEL]
y_val_dice = preprocess_val_df[LABEL]

x_test_dice = preprocess_test_df[NO_LABEL]
y_test_dice = preprocess_test_df[LABEL]

#display(x_train_dice.head())
#display(x_val_dice.head())
#display(x_test_dice.head())
```

In [92]:

```
print(len(NO_LABEL))
print(NO_LABEL)
```

```
24
['avg_score', 'sum_click_0-800', 'sum_click_1601-2400', 'sum_click_2401-3200', 'sum_c
lick_801-1600', 'sum_click_>3200', 'gender_F', 'gender_M', 'highest_education_A Level
or Equivalent', 'highest_education_HE Qualification', 'highest_education_Lower Than A
Level', 'highest_education_No Formal quals', 'highest_education_Post Graduate Qualifi
cation', 'imd_band_0-20%', 'imd_band_20-40%', 'imd_band_40-60%', 'imd_band_60-80%',
'imd_band_80-100%', 'age_band_0-35', 'age_band_35-55', 'num_of_prev_attempts_N', 'num
_of_prev_attempts_Y', 'disability_N', 'disability_Y']
```

In [93]:

```
display(x_train_dice.head())
```

	avg_score	sum_click_0-800	sum_click_1601-2400	sum_click_2401-3200	sum_click_801-1600	sum_click_>3200	g
4310	0.815000	0	0	0	1	0	
4311	0.709583	1	0	0	0	0	
4312	0.795000	0	0	1	0	0	
4313	0.748182	0	0	0	1	0	
4314	0.602000	1	0	0	0	0	

5 rows × 24 columns

In [94]:
```
print(x_train_dice.shape)
print(x_val_dice.shape)
print(x_test_dice.shape)
```

```
(7664, 24)
(1916, 24)
(2394, 24)
```

In [95]:
```
best_model_for_dice = best_model([95, 76, 95, 76])
best_model_for_dice.fit(x_train_dice.values, y_train_dice.values,
          validation_data=(x_val_dice.values, y_val_dice.values),
          epochs=100, batch_size=600,
          callbacks=[keras.callbacks.EarlyStopping(monitor="loss", patience=5)])
```

```
Epoch 1/100
13/13 [==============================] - ETA: 0s - loss: 0.6974 - accuracy: 0.4150 -
true_positives: 146.0000 - true_negatives: 103.0000 - false_positives: 72.0000 - fals
e_negatives: 279.000 - 0s 30ms/step - loss: 0.6140 - accuracy: 0.6888 - true_positive
s: 5176.0000 - true_negatives: 103.0000 - false_positives: 2103.0000 - false_negative
s: 282.0000 - val_loss: 0.5877 - val_accuracy: 0.7098 - val_true_positives: 1360.0000
- val_true_negatives: 0.0000e+00 - val_false_positives: 556.0000 - val_false_negative
s: 0.0000e+00
Epoch 2/100
13/13 [==============================] - ETA: 0s - loss: 0.5555 - accuracy: 0.7350 -
true_positives: 441.0000 - true_negatives: 0.0000e+00 - false_positives: 159.0000 - f
alse_negatives: 0.0000e+ - 0s 3ms/step - loss: 0.5592 - accuracy: 0.7122 - true_posit
ives: 5458.0000 - true_negatives: 0.0000e+00 - false_positives: 2206.0000 - false_neg
atives: 0.0000e+00 - val_loss: 0.5449 - val_accuracy: 0.7098 - val_true_positives: 13
60.0000 - val_true_negatives: 0.0000e+00 - val_false_positives: 556.0000 - val_false_
negatives: 0.0000e+00
Epoch 3/100
13/13 [==============================] - ETA: 0s - loss: 0.5413 - accuracy: 0.7067 -
true_positives: 424.0000 - true_negatives: 0.0000e+00 - false_positives: 176.0000 - f
alse_negatives: 0.0000e+ - 0s 3ms/step - loss: 0.5272 - accuracy: 0.7221 - true_posit
ives: 5292.0000 - true_negatives: 242.0000 - false_positives: 1964.0000 - false_negat
ives: 166.0000 - val_loss: 0.5340 - val_accuracy: 0.7244 - val_true_positives: 1219.0
000 - val_true_negatives: 169.0000 - val_false_positives: 387.0000 - val_false_negati
ves: 141.0000
Epoch 4/100
13/13 [==============================] - ETA: 0s - loss: 0.5195 - accuracy: 0.7383 -
true_positives: 393.0000 - true_negatives: 50.0000 - false_positives: 103.0000 - fals
e_negatives: 54.00 - 0s 2ms/step - loss: 0.5193 - accuracy: 0.7317 - true_positives:
5011.0000 - true_negatives: 597.0000 - false_positives: 1609.0000 - false_negatives:
447.0000 - val_loss: 0.5274 - val_accuracy: 0.7260 - val_true_positives: 1236.0000 -
val_true_negatives: 155.0000 - val_false_positives: 401.0000 - val_false_negatives: 1
24.0000
Epoch 5/100
13/13 [==============================] - ETA: 0s - loss: 0.5437 - accuracy: 0.7183 -
true_positives: 375.0000 - true_negatives: 56.0000 - false_positives: 128.0000 - fals
e_negatives: 41.00 - 0s 2ms/step - loss: 0.5141 - accuracy: 0.7370 - true_positives:
4943.0000 - true_negatives: 705.0000 - false_positives: 1501.0000 - false_negatives:
515.0000 - val_loss: 0.5248 - val_accuracy: 0.7286 - val_true_positives: 1227.0000 -
val_true_negatives: 169.0000 - val_false_positives: 387.0000 - val_false_negatives: 1
33.0000
Epoch 6/100
13/13 [==============================] - ETA: 0s - loss: 0.4958 - accuracy: 0.7583 -
true_positives: 391.0000 - true_negatives: 64.0000 - false_positives: 113.0000 - fals
e_negatives: 32.00 - 0s 2ms/step - loss: 0.5118 - accuracy: 0.7383 - true_positives:
4979.0000 - true_negatives: 679.0000 - false_positives: 1527.0000 - false_negatives:
479.0000 - val_loss: 0.5244 - val_accuracy: 0.7260 - val_true_positives: 1254.0000 -
val_true_negatives: 137.0000 - val_false_positives: 419.0000 - val_false_negatives: 1
06.0000
Epoch 7/100
13/13 [==============================] - ETA: 0s - loss: 0.4924 - accuracy: 0.7533 -
true_positives: 397.0000 - true_negatives: 55.0000 - false_positives: 115.0000 - fals
e_negatives: 33.00 - 0s 2ms/step - loss: 0.5088 - accuracy: 0.7397 - true_positives:
4950.0000 - true_negatives: 719.0000 - false_positives: 1487.0000 - false_negatives:
508.0000 - val_loss: 0.5266 - val_accuracy: 0.7281 - val_true_positives: 1254.0000 -
val_true_negatives: 141.0000 - val_false_positives: 415.0000 - val_false_negatives: 1
06.0000
Epoch 8/100
13/13 [==============================] - ETA: 0s - loss: 0.4804 - accuracy: 0.7517 -
true_positives: 406.0000 - true_negatives: 45.0000 - false_positives: 114.0000 - fals
e_negatives: 35.00 - 0s 2ms/step - loss: 0.5057 - accuracy: 0.7407 - true_positives:
```

```
4959.0000 - true_negatives: 718.0000 - false_positives: 1488.0000 - false_negatives: 
499.0000 - val_loss: 0.5206 - val_accuracy: 0.7291 - val_true_positives: 1224.0000 - 
val_true_negatives: 173.0000 - val_false_positives: 383.0000 - val_false_negatives: 1
36.0000
Epoch 9/100
13/13 [==============================] - ETA: 0s - loss: 0.5148 - accuracy: 0.7250 - 
true_positives: 369.0000 - true_negatives: 66.0000 - false_positives: 116.0000 - fals
e_negatives: 49.00 - 0s 2ms/step - loss: 0.5031 - accuracy: 0.7424 - true_positives: 
5039.0000 - true_negatives: 651.0000 - false_positives: 1555.0000 - false_negatives: 
419.0000 - val_loss: 0.5183 - val_accuracy: 0.7307 - val_true_positives: 1232.0000 - 
val_true_negatives: 168.0000 - val_false_positives: 388.0000 - val_false_negatives: 1
28.0000
Epoch 10/100
13/13 [==============================] - ETA: 0s - loss: 0.5311 - accuracy: 0.7233 - 
true_positives: 376.0000 - true_negatives: 58.0000 - false_positives: 130.0000 - fals
e_negatives: 36.00 - 0s 2ms/step - loss: 0.5015 - accuracy: 0.7465 - true_positives: 
4936.0000 - true_negatives: 785.0000 - false_positives: 1421.0000 - false_negatives: 
522.0000 - val_loss: 0.5149 - val_accuracy: 0.7328 - val_true_positives: 1214.0000 - 
val_true_negatives: 190.0000 - val_false_positives: 366.0000 - val_false_negatives: 1
46.0000
Epoch 11/100
13/13 [==============================] - ETA: 0s - loss: 0.5118 - accuracy: 0.7283 - 
true_positives: 376.0000 - true_negatives: 61.0000 - false_positives: 118.0000 - fals
e_negatives: 45.00 - 0s 2ms/step - loss: 0.4956 - accuracy: 0.7450 - true_positives: 
4961.0000 - true_negatives: 749.0000 - false_positives: 1457.0000 - false_negatives: 
497.0000 - val_loss: 0.5153 - val_accuracy: 0.7437 - val_true_positives: 1264.0000 - 
val_true_negatives: 161.0000 - val_false_positives: 395.0000 - val_false_negatives: 9
6.0000
Epoch 12/100
13/13 [==============================] - ETA: 0s - loss: 0.5450 - accuracy: 0.7300 - 
true_positives: 405.0000 - true_negatives: 33.0000 - false_positives: 125.0000 - fals
e_negatives: 37.00 - 0s 2ms/step - loss: 0.4903 - accuracy: 0.7508 - true_positives: 
5029.0000 - true_negatives: 725.0000 - false_positives: 1481.0000 - false_negatives: 
429.0000 - val_loss: 0.5069 - val_accuracy: 0.7484 - val_true_positives: 1243.0000 - 
val_true_negatives: 191.0000 - val_false_positives: 365.0000 - val_false_negatives: 1
17.0000
Epoch 13/100
13/13 [==============================] - ETA: 0s - loss: 0.4719 - accuracy: 0.7633 - 
true_positives: 387.0000 - true_negatives: 71.0000 - false_positives: 103.0000 - fals
e_negatives: 39.00 - 0s 2ms/step - loss: 0.4853 - accuracy: 0.7527 - true_positives: 
4904.0000 - true_negatives: 865.0000 - false_positives: 1341.0000 - false_negatives: 
554.0000 - val_loss: 0.5127 - val_accuracy: 0.7479 - val_true_positives: 1274.0000 - 
val_true_negatives: 159.0000 - val_false_positives: 397.0000 - val_false_negatives: 8
6.0000
Epoch 14/100
13/13 [==============================] - ETA: 0s - loss: 0.5025 - accuracy: 0.7550 - 
true_positives: 398.0000 - true_negatives: 55.0000 - false_positives: 122.0000 - fals
e_negatives: 25.00 - 0s 3ms/step - loss: 0.4828 - accuracy: 0.7560 - true_positives: 
4970.0000 - true_negatives: 824.0000 - false_positives: 1382.0000 - false_negatives: 
488.0000 - val_loss: 0.5013 - val_accuracy: 0.7495 - val_true_positives: 1231.0000 - 
val_true_negatives: 205.0000 - val_false_positives: 351.0000 - val_false_negatives: 1
29.0000
Epoch 15/100
13/13 [==============================] - ETA: 0s - loss: 0.4677 - accuracy: 0.7617 - 
true_positives: 381.0000 - true_negatives: 76.0000 - false_positives: 96.0000 - false
_negatives: 47.000 - 0s 2ms/step - loss: 0.4780 - accuracy: 0.7572 - true_positives: 
4900.0000 - true_negatives: 903.0000 - false_positives: 1303.0000 - false_negatives: 
558.0000 - val_loss: 0.5006 - val_accuracy: 0.7516 - val_true_positives: 1228.0000 - 
val_true_negatives: 212.0000 - val_false_positives: 344.0000 - val_false_negatives: 1
32.0000
```

```
Epoch 16/100
13/13 [==============================] - ETA: 0s - loss: 0.4740 - accuracy: 0.7467 -
true_positives: 374.0000 - true_negatives: 74.0000 - false_positives: 101.0000 - fals
e_negatives: 51.00 - 0s 2ms/step - loss: 0.4748 - accuracy: 0.7584 - true_positives:
4889.0000 - true_negatives: 923.0000 - false_positives: 1283.0000 - false_negatives:
569.0000 - val_loss: 0.5030 - val_accuracy: 0.7505 - val_true_positives: 1237.0000 -
val_true_negatives: 201.0000 - val_false_positives: 355.0000 - val_false_negatives: 1
23.0000
Epoch 17/100
13/13 [==============================] - ETA: 0s - loss: 0.4690 - accuracy: 0.7650 -
true_positives: 385.0000 - true_negatives: 74.0000 - false_positives: 95.0000 - false
_negatives: 46.000 - 0s 2ms/step - loss: 0.4730 - accuracy: 0.7608 - true_positives:
4861.0000 - true_negatives: 970.0000 - false_positives: 1236.0000 - false_negatives:
597.0000 - val_loss: 0.5019 - val_accuracy: 0.7531 - val_true_positives: 1231.0000 -
val_true_negatives: 212.0000 - val_false_positives: 344.0000 - val_false_negatives: 1
29.0000
Epoch 18/100
13/13 [==============================] - ETA: 0s - loss: 0.4764 - accuracy: 0.7583 -
true_positives: 394.0000 - true_negatives: 61.0000 - false_positives: 98.0000 - false
_negatives: 47.000 - 0s 2ms/step - loss: 0.4707 - accuracy: 0.7615 - true_positives:
4903.0000 - true_negatives: 933.0000 - false_positives: 1273.0000 - false_negatives:
555.0000 - val_loss: 0.4999 - val_accuracy: 0.7490 - val_true_positives: 1181.0000 -
val_true_negatives: 254.0000 - val_false_positives: 302.0000 - val_false_negatives: 1
79.0000
Epoch 19/100
13/13 [==============================] - ETA: 0s - loss: 0.4634 - accuracy: 0.7650 -
true_positives: 386.0000 - true_negatives: 73.0000 - false_positives: 70.0000 - false
_negatives: 71.000 - 0s 2ms/step - loss: 0.4727 - accuracy: 0.7578 - true_positives:
4789.0000 - true_negatives: 1019.0000 - false_positives: 1187.0000 - false_negatives:
669.0000 - val_loss: 0.5020 - val_accuracy: 0.7500 - val_true_positives: 1187.0000 -
val_true_negatives: 250.0000 - val_false_positives: 306.0000 - val_false_negatives: 1
73.0000
Epoch 20/100
13/13 [==============================] - ETA: 0s - loss: 0.4543 - accuracy: 0.7683 -
true_positives: 363.0000 - true_negatives: 98.0000 - false_positives: 89.0000 - false
_negatives: 50.000 - 0s 3ms/step - loss: 0.4697 - accuracy: 0.7633 - true_positives:
4861.0000 - true_negatives: 989.0000 - false_positives: 1217.0000 - false_negatives:
597.0000 - val_loss: 0.5054 - val_accuracy: 0.7537 - val_true_positives: 1251.0000 -
val_true_negatives: 193.0000 - val_false_positives: 363.0000 - val_false_negatives: 1
09.0000
Epoch 21/100
13/13 [==============================] - ETA: 0s - loss: 0.4807 - accuracy: 0.7567 -
true_positives: 386.0000 - true_negatives: 68.0000 - false_positives: 107.0000 - fals
e_negatives: 39.00 - 0s 2ms/step - loss: 0.4687 - accuracy: 0.7625 - true_positives:
4806.0000 - true_negatives: 1038.0000 - false_positives: 1168.0000 - false_negatives:
652.0000 - val_loss: 0.5132 - val_accuracy: 0.7505 - val_true_positives: 1278.0000 -
val_true_negatives: 160.0000 - val_false_positives: 396.0000 - val_false_negatives: 8
2.0000
Epoch 22/100
13/13 [==============================] - ETA: 0s - loss: 0.4369 - accuracy: 0.7750 -
true_positives: 415.0000 - true_negatives: 50.0000 - false_positives: 115.0000 - fals
e_negatives: 20.00 - 0s 2ms/step - loss: 0.4679 - accuracy: 0.7637 - true_positives:
4892.0000 - true_negatives: 961.0000 - false_positives: 1245.0000 - false_negatives:
566.0000 - val_loss: 0.5036 - val_accuracy: 0.7505 - val_true_positives: 1205.0000 -
val_true_negatives: 233.0000 - val_false_positives: 323.0000 - val_false_negatives: 1
55.0000
Epoch 23/100
13/13 [==============================] - ETA: 0s - loss: 0.4592 - accuracy: 0.7783 -
true_positives: 389.0000 - true_negatives: 78.0000 - false_positives: 89.0000 - false
_negatives: 44.000 - 0s 2ms/step - loss: 0.4640 - accuracy: 0.7649 - true_positives:
```

```
4880.0000 - true_negatives: 982.0000 - false_positives: 1224.0000 - false_negatives:
578.0000 - val_loss: 0.5033 - val_accuracy: 0.7469 - val_true_positives: 1232.0000 -
val_true_negatives: 199.0000 - val_false_positives: 357.0000 - val_false_negatives: 1
28.0000
Epoch 24/100
13/13 [==============================] - ETA: 0s - loss: 0.4852 - accuracy: 0.7667 -
true_positives: 379.0000 - true_negatives: 81.0000 - false_positives: 103.0000 - fals
e_negatives: 37.00 - 0s 2ms/step - loss: 0.4636 - accuracy: 0.7646 - true_positives:
4841.0000 - true_negatives: 1019.0000 - false_positives: 1187.0000 - false_negatives:
617.0000 - val_loss: 0.5036 - val_accuracy: 0.7516 - val_true_positives: 1181.0000 -
val_true_negatives: 259.0000 - val_false_positives: 297.0000 - val_false_negatives: 1
79.0000
Epoch 25/100
13/13 [==============================] - ETA: 0s - loss: 0.4751 - accuracy: 0.7583 -
true_positives: 370.0000 - true_negatives: 85.0000 - false_positives: 89.0000 - false
_negatives: 56.000 - 0s 2ms/step - loss: 0.4659 - accuracy: 0.7624 - true_positives:
4841.0000 - true_negatives: 1002.0000 - false_positives: 1204.0000 - false_negatives:
617.0000 - val_loss: 0.5022 - val_accuracy: 0.7500 - val_true_positives: 1219.0000 -
val_true_negatives: 218.0000 - val_false_positives: 338.0000 - val_false_negatives: 1
41.0000
Epoch 26/100
13/13 [==============================] - ETA: 0s - loss: 0.4317 - accuracy: 0.7917 -
true_positives: 389.0000 - true_negatives: 86.0000 - false_positives: 84.0000 - false
_negatives: 41.000 - 0s 2ms/step - loss: 0.4633 - accuracy: 0.7632 - true_positives:
4833.0000 - true_negatives: 1016.0000 - false_positives: 1190.0000 - false_negatives:
625.0000 - val_loss: 0.5185 - val_accuracy: 0.7505 - val_true_positives: 1277.0000 -
val_true_negatives: 161.0000 - val_false_positives: 395.0000 - val_false_negatives: 8
3.0000
Epoch 27/100
13/13 [==============================] - ETA: 0s - loss: 0.4695 - accuracy: 0.7500 -
true_positives: 390.0000 - true_negatives: 60.0000 - false_positives: 122.0000 - fals
e_negatives: 28.00 - 0s 2ms/step - loss: 0.4634 - accuracy: 0.7630 - true_positives:
4845.0000 - true_negatives: 1003.0000 - false_positives: 1203.0000 - false_negatives:
613.0000 - val_loss: 0.5070 - val_accuracy: 0.7500 - val_true_positives: 1235.0000 -
val_true_negatives: 202.0000 - val_false_positives: 354.0000 - val_false_negatives: 1
25.0000
Epoch 28/100
13/13 [==============================] - ETA: 0s - loss: 0.4574 - accuracy: 0.8033 -
true_positives: 399.0000 - true_negatives: 83.0000 - false_positives: 91.0000 - false
_negatives: 27.000 - 0s 2ms/step - loss: 0.4610 - accuracy: 0.7672 - true_positives:
4894.0000 - true_negatives: 986.0000 - false_positives: 1220.0000 - false_negatives:
564.0000 - val_loss: 0.5057 - val_accuracy: 0.7469 - val_true_positives: 1186.0000 -
val_true_negatives: 245.0000 - val_false_positives: 311.0000 - val_false_negatives: 1
74.0000
Epoch 29/100
13/13 [==============================] - ETA: 0s - loss: 0.4670 - accuracy: 0.7617 -
true_positives: 367.0000 - true_negatives: 90.0000 - false_positives: 93.0000 - false
_negatives: 50.000 - 0s 2ms/step - loss: 0.4608 - accuracy: 0.7672 - true_positives:
4817.0000 - true_negatives: 1063.0000 - false_positives: 1143.0000 - false_negatives:
641.0000 - val_loss: 0.5053 - val_accuracy: 0.7526 - val_true_positives: 1211.0000 -
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
49.0000
Epoch 30/100
13/13 [==============================] - ETA: 0s - loss: 0.4378 - accuracy: 0.7933 -
true_positives: 389.0000 - true_negatives: 87.0000 - false_positives: 80.0000 - false
_negatives: 44.000 - 0s 2ms/step - loss: 0.4599 - accuracy: 0.7684 - true_positives:
4925.0000 - true_negatives: 964.0000 - false_positives: 1242.0000 - false_negatives:
533.0000 - val_loss: 0.5035 - val_accuracy: 0.7505 - val_true_positives: 1216.0000 -
val_true_negatives: 222.0000 - val_false_positives: 334.0000 - val_false_negatives: 1
44.0000
```

```
Epoch 31/100
13/13 [==============================] - ETA: 0s - loss: 0.4242 - accuracy: 0.7883 -
true_positives: 398.0000 - true_negatives: 75.0000 - false_positives: 87.0000 - false
_negatives: 40.000 - 0s 2ms/step - loss: 0.4573 - accuracy: 0.7680 - true_positives:
4830.0000 - true_negatives: 1056.0000 - false_positives: 1150.0000 - false_negatives:
628.0000 - val_loss: 0.5088 - val_accuracy: 0.7521 - val_true_positives: 1236.0000 -
val_true_negatives: 205.0000 - val_false_positives: 351.0000 - val_false_negatives: 1
24.0000
Epoch 32/100
13/13 [==============================] - ETA: 0s - loss: 0.4754 - accuracy: 0.7517 -
true_positives: 394.0000 - true_negatives: 57.0000 - false_positives: 101.0000 - fals
e_negatives: 48.00 - 0s 2ms/step - loss: 0.4596 - accuracy: 0.7655 - true_positives:
4877.0000 - true_negatives: 990.0000 - false_positives: 1216.0000 - false_negatives:
581.0000 - val_loss: 0.5154 - val_accuracy: 0.7401 - val_true_positives: 1167.0000 -
val_true_negatives: 251.0000 - val_false_positives: 305.0000 - val_false_negatives: 1
93.0000
Epoch 33/100
13/13 [==============================] - ETA: 0s - loss: 0.4564 - accuracy: 0.7717 -
true_positives: 369.0000 - true_negatives: 94.0000 - false_positives: 74.0000 - false
_negatives: 63.000 - 0s 2ms/step - loss: 0.4621 - accuracy: 0.7697 - true_positives:
4829.0000 - true_negatives: 1070.0000 - false_positives: 1136.0000 - false_negatives:
629.0000 - val_loss: 0.5092 - val_accuracy: 0.7463 - val_true_positives: 1238.0000 -
val_true_negatives: 192.0000 - val_false_positives: 364.0000 - val_false_negatives: 1
22.0000
Epoch 34/100
13/13 [==============================] - ETA: 0s - loss: 0.4618 - accuracy: 0.7550 -
true_positives: 384.0000 - true_negatives: 69.0000 - false_positives: 108.0000 - fals
e_negatives: 39.00 - 0s 2ms/step - loss: 0.4568 - accuracy: 0.7697 - true_positives:
4879.0000 - true_negatives: 1020.0000 - false_positives: 1186.0000 - false_negatives:
579.0000 - val_loss: 0.5081 - val_accuracy: 0.7505 - val_true_positives: 1190.0000 -
val_true_negatives: 248.0000 - val_false_positives: 308.0000 - val_false_negatives: 1
70.0000
Epoch 35/100
13/13 [==============================] - ETA: 0s - loss: 0.4405 - accuracy: 0.7800 -
true_positives: 376.0000 - true_negatives: 92.0000 - false_positives: 75.0000 - false
_negatives: 57.000 - 0s 2ms/step - loss: 0.4591 - accuracy: 0.7692 - true_positives:
4903.0000 - true_negatives: 992.0000 - false_positives: 1214.0000 - false_negatives:
555.0000 - val_loss: 0.5229 - val_accuracy: 0.7516 - val_true_positives: 1262.0000 -
val_true_negatives: 178.0000 - val_false_positives: 378.0000 - val_false_negatives: 9
8.0000
Epoch 36/100
13/13 [==============================] - ETA: 0s - loss: 0.4543 - accuracy: 0.7617 -
true_positives: 398.0000 - true_negatives: 59.0000 - false_positives: 109.0000 - fals
e_negatives: 34.00 - 0s 2ms/step - loss: 0.4570 - accuracy: 0.7679 - true_positives:
4822.0000 - true_negatives: 1063.0000 - false_positives: 1143.0000 - false_negatives:
636.0000 - val_loss: 0.5045 - val_accuracy: 0.7453 - val_true_positives: 1203.0000 -
val_true_negatives: 225.0000 - val_false_positives: 331.0000 - val_false_negatives: 1
57.0000
Epoch 37/100
13/13 [==============================] - ETA: 0s - loss: 0.4509 - accuracy: 0.7850 -
true_positives: 372.0000 - true_negatives: 99.0000 - false_positives: 86.0000 - false
_negatives: 43.000 - 0s 2ms/step - loss: 0.4543 - accuracy: 0.7727 - true_positives:
4950.0000 - true_negatives: 972.0000 - false_positives: 1234.0000 - false_negatives:
508.0000 - val_loss: 0.5158 - val_accuracy: 0.7516 - val_true_positives: 1212.0000 -
val_true_negatives: 228.0000 - val_false_positives: 328.0000 - val_false_negatives: 1
48.0000
Epoch 38/100
13/13 [==============================] - ETA: 0s - loss: 0.4357 - accuracy: 0.7567 -
true_positives: 377.0000 - true_negatives: 77.0000 - false_positives: 90.0000 - false
_negatives: 56.000 - 0s 3ms/step - loss: 0.4542 - accuracy: 0.7705 - true_positives:
```

```
4824.0000 - true_negatives: 1081.0000 - false_positives: 1125.0000 - false_negatives: 
634.0000 - val_loss: 0.5183 - val_accuracy: 0.7463 - val_true_positives: 1265.0000 - 
val_true_negatives: 165.0000 - val_false_positives: 391.0000 - val_false_negatives: 9
5.0000
Epoch 39/100
13/13 [==============================] - ETA: 0s - loss: 0.4537 - accuracy: 0.7650 - 
true_positives: 410.0000 - true_negatives: 49.0000 - false_positives: 106.0000 - fals
e_negatives: 35.00 - 0s 2ms/step - loss: 0.4529 - accuracy: 0.7724 - true_positives: 
4902.0000 - true_negatives: 1018.0000 - false_positives: 1188.0000 - false_negatives: 
556.0000 - val_loss: 0.5115 - val_accuracy: 0.7396 - val_true_positives: 1159.0000 - 
val_true_negatives: 258.0000 - val_false_positives: 298.0000 - val_false_negatives: 2
01.0000
Epoch 40/100
13/13 [==============================] - ETA: 0s - loss: 0.4811 - accuracy: 0.7567 - 
true_positives: 333.0000 - true_negatives: 121.0000 - false_positives: 88.0000 - fals
e_negatives: 58.00 - 0s 2ms/step - loss: 0.4537 - accuracy: 0.7691 - true_positives: 
4820.0000 - true_negatives: 1074.0000 - false_positives: 1132.0000 - false_negatives: 
638.0000 - val_loss: 0.5144 - val_accuracy: 0.7469 - val_true_positives: 1259.0000 - 
val_true_negatives: 172.0000 - val_false_positives: 384.0000 - val_false_negatives: 1
01.0000
Epoch 41/100
13/13 [==============================] - ETA: 0s - loss: 0.4321 - accuracy: 0.7817 - 
true_positives: 408.0000 - true_negatives: 61.0000 - false_positives: 109.0000 - fals
e_negatives: 22.00 - 0s 2ms/step - loss: 0.4518 - accuracy: 0.7702 - true_positives: 
4903.0000 - true_negatives: 1000.0000 - false_positives: 1206.0000 - false_negatives: 
555.0000 - val_loss: 0.5134 - val_accuracy: 0.7547 - val_true_positives: 1215.0000 - 
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
45.0000
Epoch 42/100
13/13 [==============================] - ETA: 0s - loss: 0.4489 - accuracy: 0.7817 - 
true_positives: 401.0000 - true_negatives: 68.0000 - false_positives: 83.0000 - false
_negatives: 48.000 - 0s 2ms/step - loss: 0.4492 - accuracy: 0.7743 - true_positives: 
4869.0000 - true_negatives: 1065.0000 - false_positives: 1141.0000 - false_negatives: 
589.0000 - val_loss: 0.5141 - val_accuracy: 0.7490 - val_true_positives: 1228.0000 - 
val_true_negatives: 207.0000 - val_false_positives: 349.0000 - val_false_negatives: 1
32.0000
Epoch 43/100
13/13 [==============================] - ETA: 0s - loss: 0.4521 - accuracy: 0.7633 - 
true_positives: 374.0000 - true_negatives: 84.0000 - false_positives: 112.0000 - fals
e_negatives: 30.00 - 0s 2ms/step - loss: 0.4504 - accuracy: 0.7734 - true_positives: 
4879.0000 - true_negatives: 1048.0000 - false_positives: 1158.0000 - false_negatives: 
579.0000 - val_loss: 0.5130 - val_accuracy: 0.7479 - val_true_positives: 1200.0000 - 
val_true_negatives: 233.0000 - val_false_positives: 323.0000 - val_false_negatives: 1
60.0000
Epoch 44/100
13/13 [==============================] - ETA: 0s - loss: 0.4528 - accuracy: 0.7683 - 
true_positives: 371.0000 - true_negatives: 90.0000 - false_positives: 80.0000 - false
_negatives: 59.000 - 0s 2ms/step - loss: 0.4500 - accuracy: 0.7715 - true_positives: 
4873.0000 - true_negatives: 1040.0000 - false_positives: 1166.0000 - false_negatives: 
585.0000 - val_loss: 0.5209 - val_accuracy: 0.7510 - val_true_positives: 1257.0000 - 
val_true_negatives: 182.0000 - val_false_positives: 374.0000 - val_false_negatives: 1
03.0000
Epoch 45/100
13/13 [==============================] - ETA: 0s - loss: 0.4409 - accuracy: 0.7817 - 
true_positives: 391.0000 - true_negatives: 78.0000 - false_positives: 104.0000 - fals
e_negatives: 27.00 - 0s 2ms/step - loss: 0.4468 - accuracy: 0.7748 - true_positives: 
4844.0000 - true_negatives: 1094.0000 - false_positives: 1112.0000 - false_negatives: 
614.0000 - val_loss: 0.5120 - val_accuracy: 0.7453 - val_true_positives: 1207.0000 - 
val_true_negatives: 221.0000 - val_false_positives: 335.0000 - val_false_negatives: 1
53.0000
```

```
Epoch 46/100
13/13 [==============================] - ETA: 0s - loss: 0.4598 - accuracy: 0.7717 -
true_positives: 373.0000 - true_negatives: 90.0000 - false_positives: 88.0000 - false
_negatives: 49.000 - 0s 2ms/step - loss: 0.4463 - accuracy: 0.7753 - true_positives:
4886.0000 - true_negatives: 1056.0000 - false_positives: 1150.0000 - false_negatives:
572.0000 - val_loss: 0.5155 - val_accuracy: 0.7495 - val_true_positives: 1171.0000 -
val_true_negatives: 265.0000 - val_false_positives: 291.0000 - val_false_negatives: 1
89.0000
Epoch 47/100
13/13 [==============================] - ETA: 0s - loss: 0.4487 - accuracy: 0.7783 -
true_positives: 366.0000 - true_negatives: 101.0000 - false_positives: 75.0000 - fals
e_negatives: 58.00 - 0s 2ms/step - loss: 0.4459 - accuracy: 0.7734 - true_positives:
4880.0000 - true_negatives: 1047.0000 - false_positives: 1159.0000 - false_negatives:
578.0000 - val_loss: 0.5197 - val_accuracy: 0.7432 - val_true_positives: 1240.0000 -
val_true_negatives: 184.0000 - val_false_positives: 372.0000 - val_false_negatives: 1
20.0000
Epoch 48/100
13/13 [==============================] - ETA: 0s - loss: 0.4702 - accuracy: 0.7583 -
true_positives: 404.0000 - true_negatives: 51.0000 - false_positives: 117.0000 - fals
e_negatives: 28.00 - 0s 2ms/step - loss: 0.4467 - accuracy: 0.7791 - true_positives:
4888.0000 - true_negatives: 1083.0000 - false_positives: 1123.0000 - false_negatives:
570.0000 - val_loss: 0.5147 - val_accuracy: 0.7542 - val_true_positives: 1211.0000 -
val_true_negatives: 234.0000 - val_false_positives: 322.0000 - val_false_negatives: 1
49.0000
Epoch 49/100
13/13 [==============================] - ETA: 0s - loss: 0.4772 - accuracy: 0.7650 -
true_positives: 376.0000 - true_negatives: 83.0000 - false_positives: 87.0000 - false
_negatives: 54.000 - 0s 2ms/step - loss: 0.4430 - accuracy: 0.7782 - true_positives:
4899.0000 - true_negatives: 1065.0000 - false_positives: 1141.0000 - false_negatives:
559.0000 - val_loss: 0.5225 - val_accuracy: 0.7484 - val_true_positives: 1223.0000 -
val_true_negatives: 211.0000 - val_false_positives: 345.0000 - val_false_negatives: 1
37.0000
Epoch 50/100
13/13 [==============================] - ETA: 0s - loss: 0.4339 - accuracy: 0.7950 -
true_positives: 395.0000 - true_negatives: 82.0000 - false_positives: 86.0000 - false
_negatives: 37.000 - 0s 2ms/step - loss: 0.4460 - accuracy: 0.7758 - true_positives:
4821.0000 - true_negatives: 1125.0000 - false_positives: 1081.0000 - false_negatives:
637.0000 - val_loss: 0.5245 - val_accuracy: 0.7521 - val_true_positives: 1235.0000 -
val_true_negatives: 206.0000 - val_false_positives: 350.0000 - val_false_negatives: 1
25.0000
Epoch 51/100
13/13 [==============================] - ETA: 0s - loss: 0.4453 - accuracy: 0.7700 -
true_positives: 385.0000 - true_negatives: 77.0000 - false_positives: 103.0000 - fals
e_negatives: 35.00 - 0s 2ms/step - loss: 0.4446 - accuracy: 0.7745 - true_positives:
4923.0000 - true_negatives: 1013.0000 - false_positives: 1193.0000 - false_negatives:
535.0000 - val_loss: 0.5226 - val_accuracy: 0.7505 - val_true_positives: 1229.0000 -
val_true_negatives: 209.0000 - val_false_positives: 347.0000 - val_false_negatives: 1
31.0000
Epoch 52/100
13/13 [==============================] - ETA: 0s - loss: 0.4460 - accuracy: 0.7900 -
true_positives: 401.0000 - true_negatives: 73.0000 - false_positives: 87.0000 - false
_negatives: 39.000 - 0s 2ms/step - loss: 0.4438 - accuracy: 0.7786 - true_positives:
4884.0000 - true_negatives: 1083.0000 - false_positives: 1123.0000 - false_negatives:
574.0000 - val_loss: 0.5189 - val_accuracy: 0.7500 - val_true_positives: 1185.0000 -
val_true_negatives: 252.0000 - val_false_positives: 304.0000 - val_false_negatives: 1
75.0000
Epoch 53/100
13/13 [==============================] - ETA: 0s - loss: 0.4300 - accuracy: 0.7767 -
true_positives: 378.0000 - true_negatives: 88.0000 - false_positives: 72.0000 - false
_negatives: 62.000 - 0s 2ms/step - loss: 0.4415 - accuracy: 0.7765 - true_positives:
```

```
4886.0000 - true_negatives: 1065.0000 - false_positives: 1141.0000 - false_negatives:
572.0000 - val_loss: 0.5194 - val_accuracy: 0.7458 - val_true_positives: 1165.0000 -
val_true_negatives: 264.0000 - val_false_positives: 292.0000 - val_false_negatives: 1
95.0000
Epoch 54/100
13/13 [==============================] - ETA: 0s - loss: 0.4366 - accuracy: 0.7850 -
true_positives: 377.0000 - true_negatives: 94.0000 - false_positives: 80.0000 - false
_negatives: 49.000 - 0s 2ms/step - loss: 0.4397 - accuracy: 0.7769 - true_positives:
4858.0000 - true_negatives: 1096.0000 - false_positives: 1110.0000 - false_negatives:
600.0000 - val_loss: 0.5183 - val_accuracy: 0.7463 - val_true_positives: 1186.0000 -
val_true_negatives: 244.0000 - val_false_positives: 312.0000 - val_false_negatives: 1
74.0000
Epoch 55/100
13/13 [==============================] - ETA: 0s - loss: 0.4431 - accuracy: 0.7817 -
true_positives: 368.0000 - true_negatives: 101.0000 - false_positives: 77.0000 - fals
e_negatives: 54.00 - 0s 2ms/step - loss: 0.4415 - accuracy: 0.7779 - true_positives:
4910.0000 - true_negatives: 1052.0000 - false_positives: 1154.0000 - false_negatives:
548.0000 - val_loss: 0.5222 - val_accuracy: 0.7516 - val_true_positives: 1190.0000 -
val_true_negatives: 250.0000 - val_false_positives: 306.0000 - val_false_negatives: 1
70.0000
Epoch 56/100
13/13 [==============================] - ETA: 0s - loss: 0.4394 - accuracy: 0.7817 -
true_positives: 371.0000 - true_negatives: 98.0000 - false_positives: 77.0000 - false
_negatives: 54.000 - 0s 2ms/step - loss: 0.4388 - accuracy: 0.7818 - true_positives:
4874.0000 - true_negatives: 1118.0000 - false_positives: 1088.0000 - false_negatives:
584.0000 - val_loss: 0.5257 - val_accuracy: 0.7443 - val_true_positives: 1219.0000 -
val_true_negatives: 207.0000 - val_false_positives: 349.0000 - val_false_negatives: 1
41.0000
Epoch 57/100
13/13 [==============================] - ETA: 0s - loss: 0.4106 - accuracy: 0.8117 -
true_positives: 402.0000 - true_negatives: 85.0000 - false_positives: 77.0000 - false
_negatives: 36.000 - 0s 2ms/step - loss: 0.4382 - accuracy: 0.7782 - true_positives:
4885.0000 - true_negatives: 1079.0000 - false_positives: 1127.0000 - false_negatives:
573.0000 - val_loss: 0.5244 - val_accuracy: 0.7437 - val_true_positives: 1164.0000 -
val_true_negatives: 261.0000 - val_false_positives: 295.0000 - val_false_negatives: 1
96.0000
Epoch 58/100
13/13 [==============================] - ETA: 0s - loss: 0.4608 - accuracy: 0.7800 -
true_positives: 362.0000 - true_negatives: 106.0000 - false_positives: 73.0000 - fals
e_negatives: 59.00 - 0s 2ms/step - loss: 0.4431 - accuracy: 0.7818 - true_positives:
4836.0000 - true_negatives: 1156.0000 - false_positives: 1050.0000 - false_negatives:
622.0000 - val_loss: 0.5378 - val_accuracy: 0.7427 - val_true_positives: 1241.0000 -
val_true_negatives: 182.0000 - val_false_positives: 374.0000 - val_false_negatives: 1
19.0000
Epoch 59/100
13/13 [==============================] - ETA: 0s - loss: 0.4723 - accuracy: 0.7533 -
true_positives: 387.0000 - true_negatives: 65.0000 - false_positives: 114.0000 - fals
e_negatives: 34.00 - 0s 2ms/step - loss: 0.4438 - accuracy: 0.7751 - true_positives:
4918.0000 - true_negatives: 1022.0000 - false_positives: 1184.0000 - false_negatives:
540.0000 - val_loss: 0.5276 - val_accuracy: 0.7448 - val_true_positives: 1167.0000 -
val_true_negatives: 260.0000 - val_false_positives: 296.0000 - val_false_negatives: 1
93.0000
Epoch 60/100
13/13 [==============================] - ETA: 0s - loss: 0.4373 - accuracy: 0.7917 -
true_positives: 386.0000 - true_negatives: 89.0000 - false_positives: 67.0000 - false
_negatives: 58.000 - 0s 2ms/step - loss: 0.4392 - accuracy: 0.7775 - true_positives:
4889.0000 - true_negatives: 1070.0000 - false_positives: 1136.0000 - false_negatives:
569.0000 - val_loss: 0.5206 - val_accuracy: 0.7453 - val_true_positives: 1200.0000 -
val_true_negatives: 228.0000 - val_false_positives: 328.0000 - val_false_negatives: 1
60.0000
```

```
Epoch 61/100
13/13 [==============================] - ETA: 0s - loss: 0.4239 - accuracy: 0.7750 -
true_positives: 391.0000 - true_negatives: 74.0000 - false_positives: 86.0000 - false
_negatives: 49.000 - 0s 2ms/step - loss: 0.4367 - accuracy: 0.7820 - true_positives:
4880.0000 - true_negatives: 1113.0000 - false_positives: 1093.0000 - false_negatives:
578.0000 - val_loss: 0.5257 - val_accuracy: 0.7437 - val_true_positives: 1151.0000 -
val_true_negatives: 274.0000 - val_false_positives: 282.0000 - val_false_negatives: 2
09.0000
Epoch 62/100
13/13 [==============================] - ETA: 0s - loss: 0.4079 - accuracy: 0.8050 -
true_positives: 369.0000 - true_negatives: 114.0000 - false_positives: 66.0000 - fals
e_negatives: 51.00 - 0s 2ms/step - loss: 0.4369 - accuracy: 0.7796 - true_positives:
4886.0000 - true_negatives: 1089.0000 - false_positives: 1117.0000 - false_negatives:
572.0000 - val_loss: 0.5332 - val_accuracy: 0.7490 - val_true_positives: 1224.0000 -
val_true_negatives: 211.0000 - val_false_positives: 345.0000 - val_false_negatives: 1
36.0000
Epoch 63/100
13/13 [==============================] - ETA: 0s - loss: 0.3944 - accuracy: 0.7950 -
true_positives: 400.0000 - true_negatives: 77.0000 - false_positives: 90.0000 - false
_negatives: 33.000 - 0s 2ms/step - loss: 0.4356 - accuracy: 0.7816 - true_positives:
4873.0000 - true_negatives: 1117.0000 - false_positives: 1089.0000 - false_negatives:
585.0000 - val_loss: 0.5360 - val_accuracy: 0.7422 - val_true_positives: 1232.0000 -
val_true_negatives: 190.0000 - val_false_positives: 366.0000 - val_false_negatives: 1
28.0000
Epoch 64/100
13/13 [==============================] - ETA: 0s - loss: 0.4322 - accuracy: 0.7950 -
true_positives: 405.0000 - true_negatives: 72.0000 - false_positives: 85.0000 - false
_negatives: 38.000 - 0s 2ms/step - loss: 0.4383 - accuracy: 0.7767 - true_positives:
4890.0000 - true_negatives: 1063.0000 - false_positives: 1143.0000 - false_negatives:
568.0000 - val_loss: 0.5281 - val_accuracy: 0.7474 - val_true_positives: 1154.0000 -
val_true_negatives: 278.0000 - val_false_positives: 278.0000 - val_false_negatives: 2
06.0000
Epoch 65/100
13/13 [==============================] - ETA: 0s - loss: 0.4055 - accuracy: 0.7950 -
true_positives: 371.0000 - true_negatives: 106.0000 - false_positives: 68.0000 - fals
e_negatives: 55.00 - 0s 2ms/step - loss: 0.4340 - accuracy: 0.7828 - true_positives:
4906.0000 - true_negatives: 1093.0000 - false_positives: 1113.0000 - false_negatives:
552.0000 - val_loss: 0.5313 - val_accuracy: 0.7510 - val_true_positives: 1202.0000 -
val_true_negatives: 237.0000 - val_false_positives: 319.0000 - val_false_negatives: 1
58.0000
Epoch 66/100
13/13 [==============================] - ETA: 0s - loss: 0.3923 - accuracy: 0.7883 -
true_positives: 392.0000 - true_negatives: 81.0000 - false_positives: 77.0000 - false
_negatives: 50.000 - 0s 2ms/step - loss: 0.4313 - accuracy: 0.7846 - true_positives:
4890.0000 - true_negatives: 1123.0000 - false_positives: 1083.0000 - false_negatives:
568.0000 - val_loss: 0.5331 - val_accuracy: 0.7474 - val_true_positives: 1217.0000 -
val_true_negatives: 215.0000 - val_false_positives: 341.0000 - val_false_negatives: 1
43.0000
Epoch 67/100
13/13 [==============================] - ETA: 0s - loss: 0.4263 - accuracy: 0.7933 -
true_positives: 386.0000 - true_negatives: 90.0000 - false_positives: 79.0000 - false
_negatives: 45.000 - 0s 2ms/step - loss: 0.4322 - accuracy: 0.7844 - true_positives:
4873.0000 - true_negatives: 1139.0000 - false_positives: 1067.0000 - false_negatives:
585.0000 - val_loss: 0.5315 - val_accuracy: 0.7453 - val_true_positives: 1203.0000 -
val_true_negatives: 225.0000 - val_false_positives: 331.0000 - val_false_negatives: 1
57.0000
Epoch 68/100
13/13 [==============================] - ETA: 0s - loss: 0.4444 - accuracy: 0.7933 -
true_positives: 393.0000 - true_negatives: 83.0000 - false_positives: 90.0000 - false
_negatives: 34.000 - 0s 2ms/step - loss: 0.4310 - accuracy: 0.7860 - true_positives:
```

```
4928.0000 - true_negatives: 1096.0000 - false_positives: 1110.0000 - false_negatives:
530.0000 - val_loss: 0.5381 - val_accuracy: 0.7526 - val_true_positives: 1204.0000 -
val_true_negatives: 238.0000 - val_false_positives: 318.0000 - val_false_negatives: 1
56.0000
Epoch 69/100
13/13 [==============================] - ETA: 0s - loss: 0.4210 - accuracy: 0.7733 -
true_positives: 376.0000 - true_negatives: 88.0000 - false_positives: 90.0000 - false
_negatives: 46.000 - 0s 2ms/step - loss: 0.4306 - accuracy: 0.7811 - true_positives:
4868.0000 - true_negatives: 1118.0000 - false_positives: 1088.0000 - false_negatives:
590.0000 - val_loss: 0.5326 - val_accuracy: 0.7474 - val_true_positives: 1201.0000 -
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
59.0000
Epoch 70/100
13/13 [==============================] - ETA: 0s - loss: 0.4396 - accuracy: 0.7750 -
true_positives: 387.0000 - true_negatives: 78.0000 - false_positives: 94.0000 - false
_negatives: 41.000 - 0s 2ms/step - loss: 0.4308 - accuracy: 0.7821 - true_positives:
4899.0000 - true_negatives: 1095.0000 - false_positives: 1111.0000 - false_negatives:
559.0000 - val_loss: 0.5352 - val_accuracy: 0.7510 - val_true_positives: 1178.0000 -
val_true_negatives: 261.0000 - val_false_positives: 295.0000 - val_false_negatives: 1
82.0000
Epoch 71/100
13/13 [==============================] - ETA: 0s - loss: 0.4020 - accuracy: 0.7900 -
true_positives: 392.0000 - true_negatives: 82.0000 - false_positives: 78.0000 - false
_negatives: 48.000 - 0s 2ms/step - loss: 0.4321 - accuracy: 0.7800 - true_positives:
4887.0000 - true_negatives: 1091.0000 - false_positives: 1115.0000 - false_negatives:
571.0000 - val_loss: 0.5366 - val_accuracy: 0.7448 - val_true_positives: 1189.0000 -
val_true_negatives: 238.0000 - val_false_positives: 318.0000 - val_false_negatives: 1
71.0000
Epoch 72/100
13/13 [==============================] - ETA: 0s - loss: 0.4568 - accuracy: 0.7650 -
true_positives: 378.0000 - true_negatives: 81.0000 - false_positives: 84.0000 - false
_negatives: 57.000 - 0s 2ms/step - loss: 0.4311 - accuracy: 0.7818 - true_positives:
4858.0000 - true_negatives: 1134.0000 - false_positives: 1072.0000 - false_negatives:
600.0000 - val_loss: 0.5423 - val_accuracy: 0.7474 - val_true_positives: 1209.0000 -
val_true_negatives: 223.0000 - val_false_positives: 333.0000 - val_false_negatives: 1
51.0000
Epoch 73/100
13/13 [==============================] - ETA: 0s - loss: 0.3892 - accuracy: 0.8033 -
true_positives: 412.0000 - true_negatives: 70.0000 - false_positives: 84.0000 - false
_negatives: 34.000 - 0s 2ms/step - loss: 0.4286 - accuracy: 0.7858 - true_positives:
4909.0000 - true_negatives: 1113.0000 - false_positives: 1093.0000 - false_negatives:
549.0000 - val_loss: 0.5429 - val_accuracy: 0.7437 - val_true_positives: 1187.0000 -
val_true_negatives: 238.0000 - val_false_positives: 318.0000 - val_false_negatives: 1
73.0000
Epoch 74/100
13/13 [==============================] - ETA: 0s - loss: 0.4110 - accuracy: 0.7967 -
true_positives: 391.0000 - true_negatives: 87.0000 - false_positives: 81.0000 - false
_negatives: 41.000 - 0s 2ms/step - loss: 0.4302 - accuracy: 0.7835 - true_positives:
4891.0000 - true_negatives: 1114.0000 - false_positives: 1092.0000 - false_negatives:
567.0000 - val_loss: 0.5374 - val_accuracy: 0.7479 - val_true_positives: 1182.0000 -
val_true_negatives: 251.0000 - val_false_positives: 305.0000 - val_false_negatives: 1
78.0000
Epoch 75/100
13/13 [==============================] - ETA: 0s - loss: 0.4079 - accuracy: 0.7983 -
true_positives: 382.0000 - true_negatives: 97.0000 - false_positives: 71.0000 - false
_negatives: 50.000 - 0s 2ms/step - loss: 0.4275 - accuracy: 0.7822 - true_positives:
4885.0000 - true_negatives: 1110.0000 - false_positives: 1096.0000 - false_negatives:
573.0000 - val_loss: 0.5420 - val_accuracy: 0.7458 - val_true_positives: 1198.0000 -
val_true_negatives: 231.0000 - val_false_positives: 325.0000 - val_false_negatives: 1
62.0000
```

```
Epoch 76/100
13/13 [==============================] - ETA: 0s - loss: 0.4245 - accuracy: 0.7867 -
true_positives: 387.0000 - true_negatives: 85.0000 - false_positives: 91.0000 - false
_negatives: 37.000 - 0s 2ms/step - loss: 0.4271 - accuracy: 0.7863 - true_positives:
4863.0000 - true_negatives: 1163.0000 - false_positives: 1043.0000 - false_negatives:
595.0000 - val_loss: 0.5410 - val_accuracy: 0.7448 - val_true_positives: 1219.0000 -
val_true_negatives: 208.0000 - val_false_positives: 348.0000 - val_false_negatives: 1
41.0000
Epoch 77/100
13/13 [==============================] - ETA: 0s - loss: 0.4309 - accuracy: 0.7750 -
true_positives: 386.0000 - true_negatives: 79.0000 - false_positives: 98.0000 - false
_negatives: 37.000 - 0s 2ms/step - loss: 0.4257 - accuracy: 0.7855 - true_positives:
4930.0000 - true_negatives: 1090.0000 - false_positives: 1116.0000 - false_negatives:
528.0000 - val_loss: 0.5476 - val_accuracy: 0.7443 - val_true_positives: 1194.0000 -
val_true_negatives: 232.0000 - val_false_positives: 324.0000 - val_false_negatives: 1
66.0000
Epoch 78/100
13/13 [==============================] - ETA: 0s - loss: 0.4188 - accuracy: 0.8017 -
true_positives: 396.0000 - true_negatives: 85.0000 - false_positives: 80.0000 - false
_negatives: 39.000 - 0s 2ms/step - loss: 0.4274 - accuracy: 0.7858 - true_positives:
4890.0000 - true_negatives: 1132.0000 - false_positives: 1074.0000 - false_negatives:
568.0000 - val_loss: 0.5438 - val_accuracy: 0.7396 - val_true_positives: 1166.0000 -
val_true_negatives: 251.0000 - val_false_positives: 305.0000 - val_false_negatives: 1
94.0000
Epoch 79/100
13/13 [==============================] - ETA: 0s - loss: 0.4035 - accuracy: 0.8083 -
true_positives: 401.0000 - true_negatives: 84.0000 - false_positives: 67.0000 - false
_negatives: 48.000 - 0s 2ms/step - loss: 0.4263 - accuracy: 0.7838 - true_positives:
4892.0000 - true_negatives: 1115.0000 - false_positives: 1091.0000 - false_negatives:
566.0000 - val_loss: 0.5489 - val_accuracy: 0.7469 - val_true_positives: 1216.0000 -
val_true_negatives: 215.0000 - val_false_positives: 341.0000 - val_false_negatives: 1
44.0000
Epoch 80/100
13/13 [==============================] - ETA: 0s - loss: 0.4017 - accuracy: 0.7817 -
true_positives: 405.0000 - true_negatives: 64.0000 - false_positives: 90.0000 - false
_negatives: 41.000 - 0s 2ms/step - loss: 0.4269 - accuracy: 0.7808 - true_positives:
4917.0000 - true_negatives: 1067.0000 - false_positives: 1139.0000 - false_negatives:
541.0000 - val_loss: 0.5471 - val_accuracy: 0.7458 - val_true_positives: 1184.0000 -
val_true_negatives: 245.0000 - val_false_positives: 311.0000 - val_false_negatives: 1
76.0000
Epoch 81/100
13/13 [==============================] - ETA: 0s - loss: 0.3932 - accuracy: 0.8217 -
true_positives: 400.0000 - true_negatives: 93.0000 - false_positives: 62.0000 - false
_negatives: 45.000 - 0s 2ms/step - loss: 0.4264 - accuracy: 0.7814 - true_positives:
4801.0000 - true_negatives: 1188.0000 - false_positives: 1018.0000 - false_negatives:
657.0000 - val_loss: 0.5487 - val_accuracy: 0.7490 - val_true_positives: 1236.0000 -
val_true_negatives: 199.0000 - val_false_positives: 357.0000 - val_false_negatives: 1
24.0000
Epoch 82/100
13/13 [==============================] - ETA: 0s - loss: 0.4200 - accuracy: 0.7950 -
true_positives: 406.0000 - true_negatives: 71.0000 - false_positives: 90.0000 - false
_negatives: 33.000 - 0s 2ms/step - loss: 0.4276 - accuracy: 0.7797 - true_positives:
4910.0000 - true_negatives: 1066.0000 - false_positives: 1140.0000 - false_negatives:
548.0000 - val_loss: 0.5450 - val_accuracy: 0.7427 - val_true_positives: 1173.0000 -
val_true_negatives: 250.0000 - val_false_positives: 306.0000 - val_false_negatives: 1
87.0000
```

Out[95]:
```
<tensorflow.python.keras.callbacks.History at 0x1a1da0970c8>
```

In [96]:
```
data_dice = dice_ml.Data(dataframe=df_train_org[org_features],
```

```
                     continuous_features= ["avg_score"],
                     outcome_name=LABEL)

my_dice_model = dice_ml.Model(model=best_model_for_dice, backend='TF2')
dice = dice_ml.Dice(data_dice, my_dice_model, method="random")
```

In the following part I will make the predictions from the best model. This important part I would like to give the credit to Tuan Nguyen as he found this solution from the source code of DICE library. I have collaborate with Tuan in this part.

In [97]:
```
dice_predictions = df_test_org.copy(deep= True)

dice_predictions['final_result'] = best_model_for_dice(x_test_dice.values).numpy()

display(dice_predictions.head())
print(dice_predictions.shape)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	dis
0	61.666667	0-800	M	A Level or Equivalent	20-40%	0-35	N	
1	88.500000	1601-2400	M	No Formal quals	0-20%	0-35	N	
2	80.500000	1601-2400	M	Lower Than A Level	40-60%	0-35	N	
3	82.125000	1601-2400	F	A Level or Equivalent	20-40%	0-35	Y	
4	85.000000	0-800	M	A Level or Equivalent	80-100%	0-35	N	

```
(2394, 9)
```

In [98]:
```
def pass_or_fail(x,low,high):
    '''
        this function filter the prediction between the input low and high.
    '''
    if x >= low and x<=high:
        return 1.0
    else:
        return 0.0

dice_predictions['pred'] = dice_predictions['final_result'].apply(lambda x: pass_or_fa
```

In [99]:
```
#check if the new "pred" column has been created as expected.
dice_predictions.head()
```

Out[99]:

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	61.666667	0-800	M	A Level or Equivalent	20-40%	0-35	N	
1	88.500000	1601-2400	M	No Formal quals	0-20%	0-35	N	
2	80.500000	1601-2400	M	Lower Than A Level	40-60%	0-35	N	
3	82.125000	1601-2400	F	A Level or Equivalent	20-40%	0-35	Y	
4	85.000000	0-800	M	A Level or Equivalent	80-100%	0-35	N	

In [100…

```
#Select only rows that yield the predictions between 0.4, 0.6 from the best model
dice_pred_df = dice_predictions[dice_predictions['pred'] ==1.0]
```

In [101…

```
dice_pred_df.head()
```

Out[101]:

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	di
9	90.500000	801-1600	M	Lower Than A Level	40-60%	0-35	Y	
12	67.541667	0-800	F	Lower Than A Level	20-40%	0-35	N	
13	70.666667	0-800	F	Lower Than A Level	60-80%	0-35	N	
14	77.100000	0-800	M	A Level or Equivalent	40-60%	35-55	N	
18	73.333333	0-800	F	Lower Than A Level	40-60%	0-35	N	

In [102…

```
#Drop the label and pred columns, just keeps the features in order to predict.
dice_pred_df = dice_pred_df.drop(columns=['final_result','pred'])
```

In [103…

```
#check if the input df is ready for generate counterfactuals.
print(dice_pred_df.shape)
display(dice_pred_df.head())
```

```
(525, 8)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	di
9	90.500000	801-1600	M	Lower Than A Level	40-60%	0-35	Y	
12	67.541667	0-800	F	Lower Than A Level	20-40%	0-35	N	
13	70.666667	0-800	F	Lower Than A Level	60-80%	0-35	N	
14	77.100000	0-800	M	A Level or Equivalent	40-60%	35-55	N	
18	73.333333	0-800	F	Lower Than A Level	40-60%	0-35	N	

In [104…
```
for i in range(10):
    dice_pass = dice.generate_counterfactuals(dice_pred_df[i: i+1], total_CFs=4,desire
    dice_pass.visualize_as_dataframe(show_only_changes=True)
```

```
Diverse Counterfactuals found! total time taken: 00 min 24 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	90.5	801-1600	M	Lower Than A Level	40-60%	0-35	Y	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_atten
0	-	1601-2400	-	-	80-100%	-	
1	90.50000000000003	-	-	HE Qualification	-	-	
2	70.4	-	F	-	-	-	
3	100.0	-	-	-	-	35-55	

```
Diverse Counterfactuals found! total time taken: 00 min 31 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	67.5	0-800	F	Lower Than A Level	20-40%	0-35	N	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	-	2401-3200	-	Post Graduate Qualification	-	-	-	
1	51.2	801-1600	-	-	60-80%	-	-	
2	-	>3200	M	-	-	-	-	
3	93.5	-	-	-	-	35-55	-	

```
Diverse Counterfactuals found! total time taken: 00 min 24 sec
Query instance (original outcome : 1)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	70.7	0-800	F	Lower Than A Level	60-80%	0-35	N	

```
Diverse Counterfactual set (new outcome: 0.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_atten
0	70.69999999999999	-	M	-	0-20%	-	
1	43.7	-	-	-	-	-	
2	94.8	-	-	No Formal quals	-	-	
3	70.69999999999997	-	-	-	-	35-55	

```
Diverse Counterfactuals found! total time taken: 00 min 21 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	77.1	0-800	M	A Level or Equivalent	40-60%	35-55	N	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	28.6	-	-	-	-	-	-	
1	-	2401-3200	-	Post Graduate Qualification	-	-	-	
2	100.0	-	F	-	-	-	-	
3	-	-	-	-	80-100%	0-35	-	

```
Diverse Counterfactuals found! total time taken: 00 min 26 sec
Query instance (original outcome : 1)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	73.3	0-800	F	Lower Than A Level	40-60%	0-35	N	

```
Diverse Counterfactual set (new outcome: 0.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_atten
0	73.30000000000003	-	-	No Formal quals	60-80%	-	
1	46.3	2401-3200	-	-	-	-	
2	84.7	-	M	No Formal quals	-	35-55	
3	73.30000000000001	-	-	-	-	-	

```
Diverse Counterfactuals found! total time taken: 00 min 24 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	69.9	0-800	F	HE Qualification	80-100%	35-55	Y	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_atten
0	43.0	2401-3200	-	-	-	-	
1	69.90000000000003	1601-2400	M	-	-	0-35	
2	69.89999999999998	-	-	Post Graduate Qualification	-	-	
3	100.0	-	-	-	-	-	

```
Diverse Counterfactuals found! total time taken: 00 min 36 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	56.0	0-800	F	Lower Than A Level	60-80%	0-35	N	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_atte
0	44.9	1601-2400	-	A Level or Equivalent	-	-	
1	56.00000000000001	2401-3200	-	-	-	35-55	
2	90.6	-	-	-	-	-	
3	55.999999999999986	>3200	-	-	80-100%	-	

```
Diverse Counterfactuals found! total time taken: 00 min 29 sec
Query instance (original outcome : 1)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	86.0	0-800	F	Lower Than A Level	0-20%	35-55	N	

```
Diverse Counterfactual set (new outcome: 0.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attem
0	26.0	-	-	-	-	-	
1	80.1	801-1600	-	Post Graduate Qualification	40-60%	-	
2	86.00000000000004	-	-	-	-	-	
3	99.2	-	-	No Formal quals	-	-	

```
Diverse Counterfactuals found! total time taken: 00 min 43 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	75.0	0-800	M	HE Qualification	20-40%	0-35	N	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attem
0	75.00000000000001	-	F	-	-	-	
1	99.3	801-1600	-	A Level or Equivalent	-	-	
2	75.00000000000003	>3200	-	Post Graduate Qualification	80-100%	-	
3	65.2	1601-2400	-	-	-	-	

```
Diverse Counterfactuals found! total time taken: 00 min 47 sec
Query instance (original outcome : 0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	0.0	801-1600	F	A Level or Equivalent	60-80%	35-55	N	

```
Diverse Counterfactual set (new outcome: 1.0)
```

	avg_score	sum_click	gender	highest_education	imd_band	age_band	num_of_prev_attempts	disa
0	-	-	-	-	40-60%	0-35	Y	
1	-	-	-	No Formal quals	-	-	-	
2	-	0-800	M	-	-	-	-	
3	49.8	-	-	-	-	-	-	

In case 1,2,3,5, the counterfactuals points out that the outcome of the student can be changed by increasing sum_click

In case 4, the counterfactuals points out that the outcome of the student can be changed by improving avg_score

From the above counterfactual, I think avg_score and sum_click are the most important factors that a student can change to improve their success. In this part I also want to give credit to Tuan Nguyen, who found that in order to work with Dice, the org_test_df must have the same order of features like dice_pass.visualize_as_dataframe.

In [ ]: