In [12]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [38]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\C9_Data - C9_Data.csv")
df
```

Out[38]:

	row_id	user_id	timestamp	gate_id
0	0	18	2022-07-29 09:08:54	7
1	1	18	2022-07-29 09:09:54	9
2	2	18	2022-07-29 09:09:54	9
3	3	18	2022-07-29 09:10:06	5
4	4	18	2022-07-29 09:10:08	5
...	...	...	...	...
37513	37513	6	2022-12-31 20:38:56	11
37514	37514	6	2022-12-31 20:39:22	6
37515	37515	6	2022-12-31 20:39:23	6
37516	37516	6	2022-12-31 20:39:31	9
37517	37517	6	2022-12-31 20:39:31	9

37518 rows × 4 columns

In [39]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37518 entries, 0 to 37517
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   row_id     37518 non-null  int64 
 1   user_id    37518 non-null  int64 
 2   timestamp  37518 non-null  object
 3   gate_id    37518 non-null  int64 
dtypes: int64(3), object(1)
memory usage: 1.1+ MB
```

In [41]:
```
df['gate_id'].value_counts()
```

Out[41]:
```
 4     8170
 3     5351
 10    4767
 5     4619
 11    4090
 9     3390
 7     3026
 6     1800
 13    1201
 12     698
 15     298
-1       48
 8       48
 1        5
 16       4
 0        2
 14       1
Name: gate_id, dtype: int64
```

In [42]:
```
df.columns
```

Out[42]:
```
Index(['row_id', 'user_id', 'timestamp', 'gate_id'], dtype='object')
```

In [43]:
```
df1=df[['row_id','user_id','timestamp','gate_id']]
```

In [44]:
```
x=df1[['row_id', 'user_id']]
y=df['gate_id']
```

In [45]:
```
g1={'gate_id':{'15':0,'14':1,'13':2,'12':3,'11':4,'10':5,'9':6,'8':7,'7':8,'6'
:12,'1':13,'0':14,'-1':15}}
df1=df1.replace(g1)
print(df1)
```

```
       row_id  user_id            timestamp  gate_id
0           0       18  2022-07-29 09:08:54        7
1           1       18  2022-07-29 09:09:54        9
2           2       18  2022-07-29 09:09:54        9
3           3       18  2022-07-29 09:10:06        5
4           4       18  2022-07-29 09:10:08        5
...       ...      ...                  ...      ...
37513   37513        6  2022-12-31 20:38:56       11
37514   37514        6  2022-12-31 20:39:22        6
37515   37515        6  2022-12-31 20:39:23        6
37516   37516        6  2022-12-31 20:39:31        9
37517   37517        6  2022-12-31 20:39:31        9

[37518 rows x 4 columns]
```

In [46]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30)
```

In [47]:
```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[47]:
```
RandomForestClassifier()
```

In [48]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
            'n_estimators':[10,20,30,40,50]}
```

In [49]:
```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring='acc
grid_search.fit(x_train,y_train)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\model_selection\_split.py:
666: UserWarning: The least populated class in y has only 1 members, which is
less than n_splits=2.
  warnings.warn(("The least populated class in y has only %d"
```

Out[49]:
```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 10, 15, 20, 25],
                         'n_estimators': [10, 20, 30, 40, 50]},
             scoring='accuracy')
```

In [50]:
```
grid_search.best_score_
```

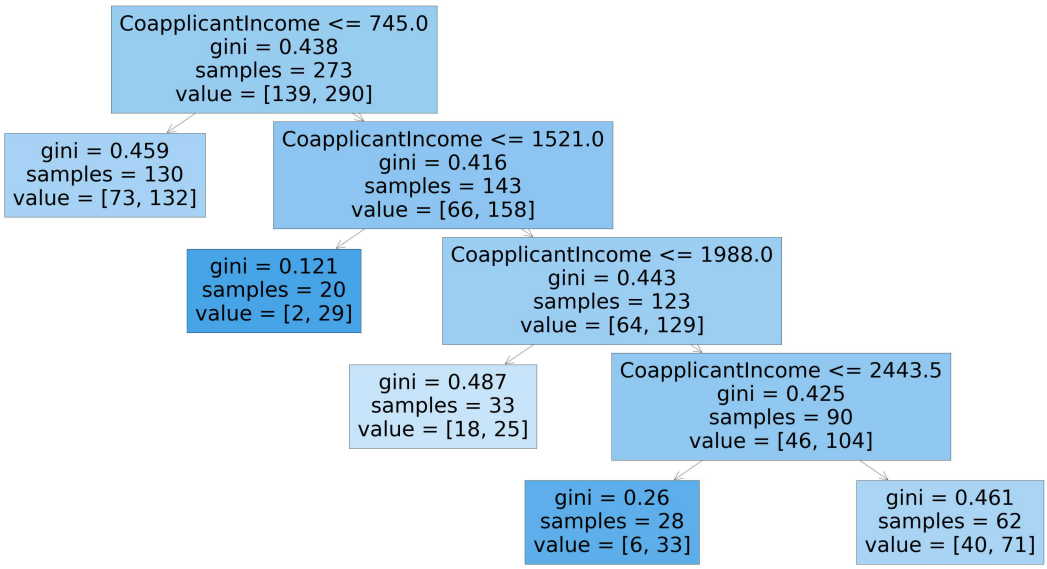
Out[50]:
```
0.22465920341177364
```

In [51]:
```
rfc_best=grid_search.best_estimator_
```

In [37]:
```
from sklearn.tree import plot_tree

plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,filled=True)
```

Out[37]:
```
[Text(1275.4285714285713, 1956.96, 'CoapplicantIncome <= 745.0\ngini = 0.438
\nsamples = 273\nvalue = [139, 290]'),
 Text(637.7142857142857, 1522.0800000000002, 'gini = 0.459\nsamples = 130\nva
lue = [73, 132]'),
 Text(1913.1428571428569, 1522.0800000000002, 'CoapplicantIncome <= 1521.0\ng
ini = 0.416\nsamples = 143\nvalue = [66, 158]'),
 Text(1275.4285714285713, 1087.2, 'gini = 0.121\nsamples = 20\nvalue = [2, 2
9]'),
 Text(2550.8571428571427, 1087.2, 'CoapplicantIncome <= 1988.0\ngini = 0.443
\nsamples = 123\nvalue = [64, 129]'),
 Text(1913.1428571428569, 652.3200000000002, 'gini = 0.487\nsamples = 33\nval
ue = [18, 25]'),
 Text(3188.5714285714284, 652.3200000000002, 'CoapplicantIncome <= 2443.5\ngi
ni = 0.425\nsamples = 90\nvalue = [46, 104]'),
 Text(2550.8571428571427, 217.44000000000005, 'gini = 0.26\nsamples = 28\nval
ue = [6, 33]'),
 Text(3826.2857142857138, 217.44000000000005, 'gini = 0.461\nsamples = 62\nva
lue = [40, 71]')]
```



In [ ]:

In [ ]: