In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
df=pd.read_csv('data.csv')
df
```

	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

In [3]:

```
df.columns
```

Out[3]:

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

In [15]:

```
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 [17]:

```
df['gate_id'].value_counts()
```

Out[17]:

```
 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 [16]:

```
x=df[['row_id', 'user_id']]
y=df['gate_id']
```

In [18]:

```
d={"gate_id":{'4':1,'3':2,'10':3,'5':4,'11':5,'9':45,'7':45,'6':6,'13':12,'12':13,'15':24,'-1':34,'8':8,'1':24,'16':32,'0':221,'14':345}
df=df.replace(df)
print(df)
```

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

[37518 rows x 4 columns]
```

In [19]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.70)
```

In [20]:

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[20]:

```
RandomForestClassifier()
```

Depth of Tree

In [21]:

```
parameters={"max_depth":[1,2,3,4,5],"min_samples_leaf":[5,23,45,76,78],'n_estimators':[10,23,45,65,7]}
```

Cross Validate

In [22]:

```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

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

Out[22]:

```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 23, 45, 76, 78],
                         'n_estimators': [10, 23, 45, 65, 7]},
             scoring='accuracy')
```

Score

In [23]:

```
grid_search.best_score_
```

Out[23]:

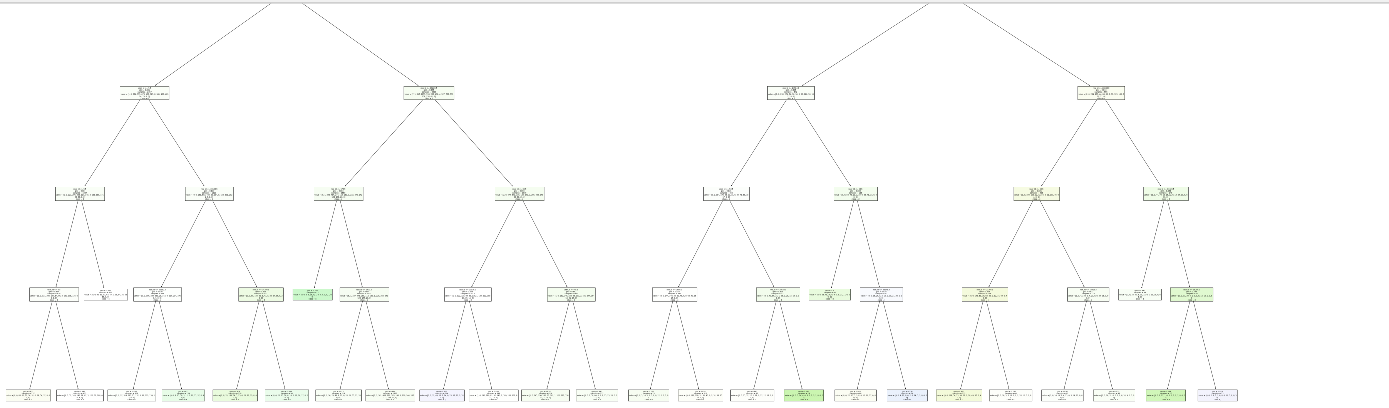
```
0.2247001271537158
```

In [24]:

```
rfc_best=grid_search.best_estimator_
```

In [26]:

```
ee import plot_tree
ize=(80,40))
est.estimators_[5],feature_names=x.columns,class_names=['a','b','c','d','e','f','g','h','i','j','k','l','m','n','o','p','q'],filled=True)
```



In [ ]: