

In [1]:

```
import np
import pandas as pd
```

In [2]:

```
X = np.load('X_train_task1.npy', mmap_mode='r')
Y = np.load('y_train_task1.npy', mmap_mode='r')

X_test_task1 = np.load('X_test_task1.npy', mmap_mode='r')
#X_test_task1 = np.array(X_test_task1)
```

## Modelling

In [5]:

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test=train_test_split(X,Y) #,test_size=0.3,random_state=
```

## Random Forest Classification

In [6]:

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import f1_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
```

In [7]:

```
rf=RandomForestClassifier(n_estimators=100,random_state=1)
rf.fit(x_train,y_train)
```

Out[7]:

```
RandomForestClassifier(random_state=1)
```

In [8]:

```
rf_pred=rf.predict(x_test)
f1_rf=f1_score(y_test,rf_pred)
precision_rf = precision_score(y_test, rf_pred)
recall_rf = recall_score(y_test, rf_pred)

print("RF f1 score: ",f1_rf)
print("RF Precision: ",precision_rf)
print("RF Recall:",recall_rf)
print("RF accuracy score: ",rf.score(x_test,y_test))
```

```
RF f1 score: 0.9905660377358491
RF Precision: 0.9813084112149533
RF Recall: 1.0
RF accuracy score: 0.9916666666666667
```

In [9]:

```
pd.DataFrame({"Id": np.arange(len(Y)), "Category": Y}).astype(int).to_csv(
    "solution.csv", index=False
)
```

In [ ]: