```python
import pandas as pd
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

def preprocess(df):
    df['Fare'] = df['Fare'].fillna(df['Fare'].mean())
    df['Age'] = df['Age'].fillna(df['Age'].mean())
    df['Embarked'] = df['Embarked'].fillna('Unknown')
    df['Sex'] = df['Sex'].apply(lambda x: 1 if x == 'male' else 0)
    df['Embarked'] = df['Embarked'].map( {'S': 0, 'C': 1, 'Q': 2, 'Unknown': 3} ).asty
pe(int)
    df = df.drop(['Cabin','Name','PassengerId','Ticket'],axis=1)
    return df

def train(df):
    train_x = df.drop('Survived', axis=1)
    train_y = df.Survived
    (train_x, test_x ,train_y, test_y) = train_test_split(train_x, train_y, test_size
 = 0.6, random_state = 42)
    dtrain = xgb.DMatrix(train_x, label=train_y)
    param = {'max_depth':3, 'learning_rate': 0.6, 'objective':'binary:logistic' }
    num_round = 2
    bst = xgb.train(param, dtrain, num_round)
    preds = bst.predict(xgb.DMatrix(test_x))
    print(accuracy_score(preds.round(), test_y))

    return bst

def predict(bst, df):
    return bst.predict(xgb.DMatrix(df))

df = pd.read_csv('./train.csv')
df_test_origin = pd.read_csv('./test.csv')
df = preprocess(df)
df_test = preprocess(df_test_origin)
bst = train(df)
answer = predict(bst, df_test).round().astype(int)
submit_data =  pd.Series(answer, name='Survived', index=df_test_origin['PassengerId'])
submit_data.to_csv('submit2.csv', header=True)
```

精度は 0.78 ほど。

| 4493 | new | Taiki Sakamoto | | 0.78468 | 2 | now |