

```
from google.colab import files
uploaded = files.upload()
```

Choose Files train.csv

train.csv(text/csv) - 61194 bytes, last modified: 4/15/2026 - 100% done
Saving train.csv to train (2).csv

```
# 1) استيراد المكتبات
import pandas as pd
import numpy as np

# 2) رفع الملف
from google.colab import files
uploaded = files.upload()

# 3) قراءة البيانات
df = pd.read_csv('train.csv')

# عرض أول 5 صفوف
print(df.head())

# -----

# 4) معالجة القيم الناقصة
print(df.isnull().sum())

# تعبئة القيم الرقمية بالمتوسط
df = df.fillna(df.mean(numeric_only=True))

# تعبئة النصوص
df = df.fillna(method='ffill')

# -----

# 5) تحويل البيانات النصية إلى أرقام
df = pd.get_dummies(df, drop_first=True)

# -----

# 6) تحديد الهدف (target)
# الهدف غالبًا اسمه Survived في Titanic
X = df.drop('Survived', axis=1)
y = df['Survived']

# -----

# 7) تقسيم البيانات
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

```
# -----

# 8) تدريب موديل + Grid Search
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV

model = RandomForestClassifier()

params = {
    'n_estimators': [50, 100],
    'max_depth': [3, 5, 10]
}

grid = GridSearchCV(model, params, cv=3)
grid.fit(X_train, y_train)

print("Best Params:", grid.best_params_)

# -----

# 9) التقييم
from sklearn.metrics import accuracy_score

y_pred = grid.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))

# -----

# 10) Feature Selection (طريقتين)

# الطريقة 1: SelectKBest
from sklearn.feature_selection import SelectKBest, chi2

selector = SelectKBest(score_func=chi2, k=5)
X_new = selector.fit_transform(X, y)

# الطريقة 2: باستخدام RandomForest
model.fit(X, y)
importances = model.feature_importances_

print("Feature Importances:", importances)

# -----

# 11) تدريب مرة ثانية بالميزات المختارة

X_train, X_test, y_train, y_test = train_test_split(
    X_new, y, test_size=0.2, random_state=42
)

model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print("New Accuracy:", accuracy_score(y_test, y_pred))
```

Choose Files train.csv

train.csv(text/csv) - 61194 bytes, last modified: 4/15/2026 - 100% done

Saving train.csv to train (3).csv

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

	Name	Sex	Age	Si
0	Braund, Mr. Owen Harris	male	22.0	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	
2	Heikkinen, Miss. Laina	female	26.0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	
4	Allen, Mr. William Henry	male	35.0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

/tmp/ipykernel_55206/2546264132.py:24: FutureWarning: DataFrame.fillna
df = df.fillna(method='ffill')

Best Params: {'max_depth': 10, 'n_estimators': 50}

Accuracy: 0.6927374301675978

Feature Importances: [0.0382034 0.05117937 0.04465976 ... 0.00018286]

New Accuracy: 0.7988826815642458

