```
# 1.Opracować przepływ pracy uczenia maszynowego zagadnienia
klasyfikacji (pojedyncze drzewo decyzyjne)
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, export_text
from sklearn.metrics import classification_report, accuracy_score

data = pd.read_csv('smokers.csv')
data.dropna()

data['Smoking Status'] = data['Smoking Status'].map({'current': 1,
'former': 0, 'never': -1})
data['Gender'] = data['Gender'].map({'f': 0, 'm': 1})

X = data.drop(columns=['GSM', 'Smoking Status'])
y = data['Smoking Status']

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

clf = DecisionTreeClassifier(random_state=42)
clf.fit(X_train, y_train)

y_pred = clf.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test,
y_pred))
print("\nDecision Tree:\n")
print(export_text(clf, feature_names=list(X.columns)))
```

```
Accuracy: 0.656934306569343

Classification Report:
               precision    recall  f1-score   support

          -1       0.39      0.36      0.37        39
           1       0.75      0.78      0.76        98

    accuracy                           0.66       137
   macro avg       0.57      0.57      0.57       137
weighted avg       0.65      0.66      0.65       137


Decision Tree:

|--- cg02839557 <= 0.04
|   |--- Age <= 32.00
|   |   |--- class: -1
|   |--- Age >  32.00
|   |   |--- cg00213748 <= 0.92
|   |   |   |--- cg03052502 <= 0.99
|   |   |   |   |--- class: 1
|   |   |   |--- cg03052502 >  0.99
|   |   |   |   |--- class: -1
|   |   |--- cg00213748 >  0.92
|   |   |   |--- class: -1
|--- cg02839557 >  0.04
|   |--- cg01707559 <= 0.06
|   |   |--- cg03695421 <= 0.58
|   |   |   |--- class: 1
|   |   |--- cg03695421 >  0.58
|   |   |   |--- class: -1
|   |--- cg01707559 >  0.06
|   |   |--- cg00050873 <= 0.57
|   |   |   |--- cg01707559 <= 0.22
|   |   |   |   |--- cg03052502 <= 0.45
|   |   |   |   |   |--- Age <= 62.50
|   |   |   |   |   |   |--- cg03443143 <= 0.38
|   |   |   |   |   |   |   |--- cg02004872 <= 0.26
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg02004872 >  0.26
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- cg03443143 >  0.38
|   |   |   |   |   |   |   |--- Age <= 48.50
|   |   |   |   |   |   |   |   |--- cg02842889 <= 0.39
|   |   |   |   |   |   |   |   |   |--- cg01707559 <= 0.22
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- cg01707559 >  0.22
|   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |--- cg02842889 >  0.39
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- Age >  48.50
|   |   |   |   |   |   |   |   |--- cg02004872 <= 0.10
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |--- cg02004872 >  0.10
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- Age >  62.50
|   |   |   |   |   |   |--- cg02494853 <= 0.08
|   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |--- cg02494853 >  0.08
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- cg03052502 >  0.45
|   |   |   |   |   |--- cg03155755 <= 0.45
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- cg03155755 >  0.45
|   |   |   |   |   |   |--- cg03683899 <= 0.33
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- cg03683899 >  0.33
|   |   |   |   |   |   |   |--- class: -1
|   |   |   |--- cg01707559 >  0.22
|   |   |   |   |--- cg00455876 <= 0.35
|   |   |   |   |   |--- cg03695421 <= 0.27
|   |   |   |   |   |   |--- cg03155755 <= 0.26
|   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |--- cg03155755 >  0.26
|   |   |   |   |   |   |   |--- cg02004872 <= 0.12
|   |   |   |   |   |   |   |   |--- cg02004872 <= 0.10
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- cg02004872 >  0.10
|   |   |   |   |   |   |   |   |   |--- cg02494853 <= 0.03
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- cg02494853 >  0.03
|   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg02004872 >  0.12
|   |   |   |   |   |   |   |   |--- cg00213748 <= 0.10
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |--- cg00213748 >  0.10
|   |   |   |   |   |   |   |   |   |--- cg00214611 <= 0.30
|   |   |   |   |   |   |   |   |   |   |--- cg00050873 <= 0.46
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- cg00050873 >  0.46
|   |   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |--- cg00214611 >  0.30
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- cg03695421 >  0.27
|   |   |   |   |   |   |--- cg02842889 <= 0.39
|   |   |   |   |   |   |   |--- cg02233190 <= 0.46
|   |   |   |   |   |   |   |   |--- Age <= 31.00
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |--- Age >  31.00
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- cg02233190 >  0.46
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |--- cg02842889 >  0.39
|   |   |   |   |   |   |   |--- cg03695421 <= 0.28
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg03695421 >  0.28
|   |   |   |   |   |   |   |   |--- cg03052502 <= 0.52
|   |   |   |   |   |   |   |   |   |--- cg03052502 <= 0.45
|   |   |   |   |   |   |   |   |   |   |--- cg02842889 <= 0.42
|   |   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |   |--- cg02842889 >  0.42
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of
depth 4
|   |   |   |   |   |   |   |   |   |--- cg03052502 >  0.45
|   |   |   |   |   |   |   |   |   |   |--- cg01707559 <= 0.28
|   |   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |   |--- cg01707559 >  0.28
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of
```

```
depth 2
|   |   |   |   |   |   |   |   |--- cg03052502 >  0.52
|   |   |   |   |   |   |   |   |   |--- cg03244189 <= 0.35
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- cg03244189 >  0.35
|   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |--- cg00455876 >  0.35
|   |   |   |   |   |--- cg00455876 <= 0.37
|   |   |   |   |   |   |--- cg02842889 <= 0.40
|   |   |   |   |   |   |   |--- cg03695421 <= 0.21
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg03695421 >  0.21
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- cg02842889 >  0.40
|   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |--- cg00455876 >  0.37
|   |   |   |   |   |   |--- cg02494853 <= 0.09
|   |   |   |   |   |   |   |--- cg00212031 <= 0.58
|   |   |   |   |   |   |   |   |--- cg03706273 <= 0.13
|   |   |   |   |   |   |   |   |   |--- cg02494853 <= 0.06
|   |   |   |   |   |   |   |   |   |   |--- cg02842889 <= 0.33
|   |   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |   |--- cg02842889 >  0.33
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of
depth 8
|   |   |   |   |   |   |   |   |   |--- cg02494853 >  0.06
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- cg03706273 >  0.13
|   |   |   |   |   |   |   |   |   |--- cg02842889 <= 0.49
|   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |--- cg02842889 >  0.49
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- cg00212031 >  0.58
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |--- cg02494853 >  0.09
|   |   |   |   |   |   |   |--- class: -1
|   |   |--- cg00050873 >  0.57
|   |   |   |--- cg02233190 <= 0.02
|   |   |   |   |--- class: -1
|   |   |   |--- cg02233190 >  0.02
|   |   |   |   |--- cg02494853 <= 0.11
|   |   |   |   |   |--- Age <= 48.50
|   |   |   |   |   |   |--- cg03695421 <= 0.64
|   |   |   |   |   |   |   |--- cg02494853 <= 0.05
|   |   |   |   |   |   |   |   |--- cg03052502 <= 0.98
|   |   |   |   |   |   |   |   |   |--- cg02494853 <= 0.03
|   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |--- cg02494853 >  0.03
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- cg03052502 >  0.98
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg02494853 >  0.05
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |--- cg03695421 >  0.64
|   |   |   |   |   |   |   |--- cg03244189 <= 0.11
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg03244189 >  0.11
|   |   |   |   |   |   |   |   |--- cg02842889 <= 0.05
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- cg02842889 >  0.05
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |--- Age >  48.50
|   |   |   |   |   |   |--- cg03695421 <= 0.20
|   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |--- cg03695421 >  0.20
|   |   |   |   |   |   |   |--- cg02839557 <= 0.04
|   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |--- cg02839557 >  0.04
|   |   |   |   |   |   |   |   |--- cg00212031 <= 0.02
|   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |--- cg00212031 >  0.02
|   |   |   |   |   |   |   |   |   |--- Age <= 67.50
|   |   |   |   |   |   |   |   |   |   |--- cg03052502 <= 0.29
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of
depth 2
|   |   |   |   |   |   |   |   |   |   |--- cg03052502 >  0.29
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of
depth 5
|   |   |   |   |   |   |   |   |   |--- Age >  67.50
|   |   |   |   |   |   |   |   |   |   |--- cg02004872 <= 0.02
|   |   |   |   |   |   |   |   |   |   |   |--- class: -1
|   |   |   |   |   |   |   |   |   |   |--- cg02004872 >  0.02
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- cg02494853 >  0.11
|   |   |   |   |   |--- class: -1


# 2. wykonać klasyfikację ensemble (używając modeli Random Forrest,
Boosting, Bagging, Gradient Boosting)
from sklearn.impute import SimpleImputer
from sklearn.ensemble import RandomForestClassifier,
BaggingClassifier, AdaBoostClassifier, GradientBoostingClassifier
from sklearn.metrics import classification_report, accuracy_score
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=42)

imputer = SimpleImputer(strategy='mean')
```

```
X_train = imputer.fit_transform(X_train)
X_test = imputer.transform(X_test)

models = {
    "Random Forest": RandomForestClassifier(n_estimators=100,
random_state=42),
    "Bagging": BaggingClassifier(n_estimators=100, random_state=42),
    "Boosting (AdaBoost)": AdaBoostClassifier(n_estimators=100,
random_state=42),
    "Gradient Boosting": GradientBoostingClassifier(n_estimators=100,
random_state=42)
}

for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    print(f"=== {name} ===")
    print(f"Accuracy: {accuracy_score(y_test, y_pred):.2f}")
    print(classification_report(y_test, y_pred))
```

```
=== Random Forest ===
Accuracy: 0.74
              precision    recall  f1-score   support

          -1       0.50      0.21      0.29        53
           1       0.77      0.93      0.84       152

    accuracy                           0.74       205
   macro avg       0.64      0.57      0.57       205
weighted avg       0.70      0.74      0.70       205


c:\Programs\Pythonek\3.11\Lib\site-packages\sklearn\impute\
_base.py:635: UserWarning: Skipping features without any observed
values: ['Gender']. At least one non-missing value is needed for
imputation with strategy='most_frequent'.
  warnings.warn(
c:\Programs\Pythonek\3.11\Lib\site-packages\sklearn\impute\
_base.py:635: UserWarning: Skipping features without any observed
values: ['Gender']. At least one non-missing value is needed for
imputation with strategy='most_frequent'.
  warnings.warn(

=== Bagging ===
Accuracy: 0.74
              precision    recall  f1-score   support

          -1       0.50      0.32      0.39        53
           1       0.79      0.89      0.84       152

    accuracy                           0.74       205
   macro avg       0.64      0.60      0.61       205
weighted avg       0.71      0.74      0.72       205

=== Boosting (AdaBoost) ===
Accuracy: 0.72
              precision    recall  f1-score   support

          -1       0.43      0.28      0.34        53
           1       0.78      0.87      0.82       152

    accuracy                           0.72       205
   macro avg       0.60      0.58      0.58       205
weighted avg       0.69      0.72      0.70       205

=== Gradient Boosting ===
Accuracy: 0.71
              precision    recall  f1-score   support

          -1       0.39      0.23      0.29        53
           1       0.76      0.88      0.82       152

    accuracy                           0.71       205
   macro avg       0.58      0.55      0.55       205
weighted avg       0.67      0.71      0.68       205
```