Dataset: bank-tr.csv

Experiment: exp_01

DecisionTreeClassifier(random_state=0)

DT experiment best results:

Experiment best score (accuracy): 0.684798

DT cross validation scores:
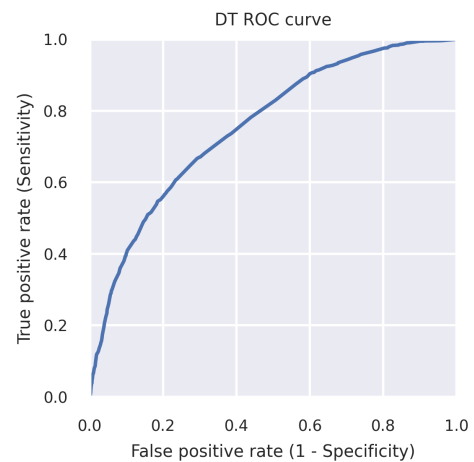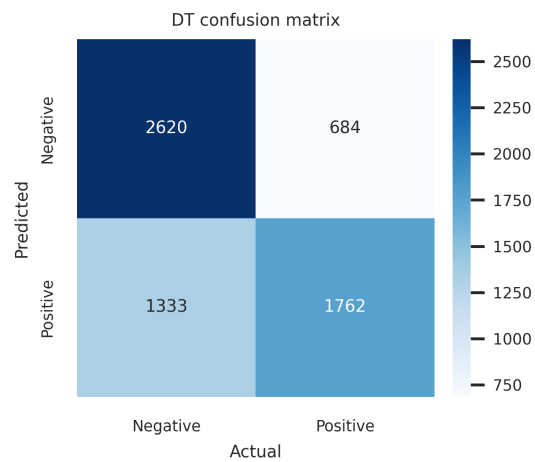
Accuracy: 0.684798
Precision: 0.723145
Recall: 0.569305
F1: 0.634849
AUC: 0.754097

DT confusion matrix & ROC curve:

Best classifier:

DecisionTreeClassifier(random_state=0)

Best hyperparameters:

ccp_alpha: 0.001
criterion: entropy
max_depth: 9
max_features: sqrt

Grid search hyperparameters:

ccp_alpha: [0.1, 0.01, 0.001]
criterion: ['gini', 'entropy']
max_depth: [5, 6, 7, 8, 9]
max_features: ['sqrt', 'log2']

Dataset: bank-tr.csv

Experiment: exp_01

DecisionTreeClassifier(random_state=0)

Experiment parameters:
    n_splits: 5
    scoring: accuracy
    target: made_deposit
    categorical columns:
        town
        country
        job
        married
        education
        defaulted?
        housing
        has_loan
        last_contact
        cc_tr
        last_contact_month
        poutcome
    feature selection:
        accountID: False
        town: True
        country: True
        age: True
        job: True
        married: True
        education: True
        defaulted?: True
        current_balance: True
        housing: True
        has_loan: True
        last_contact: True
        cc_tr: True
        last_contact_day: True
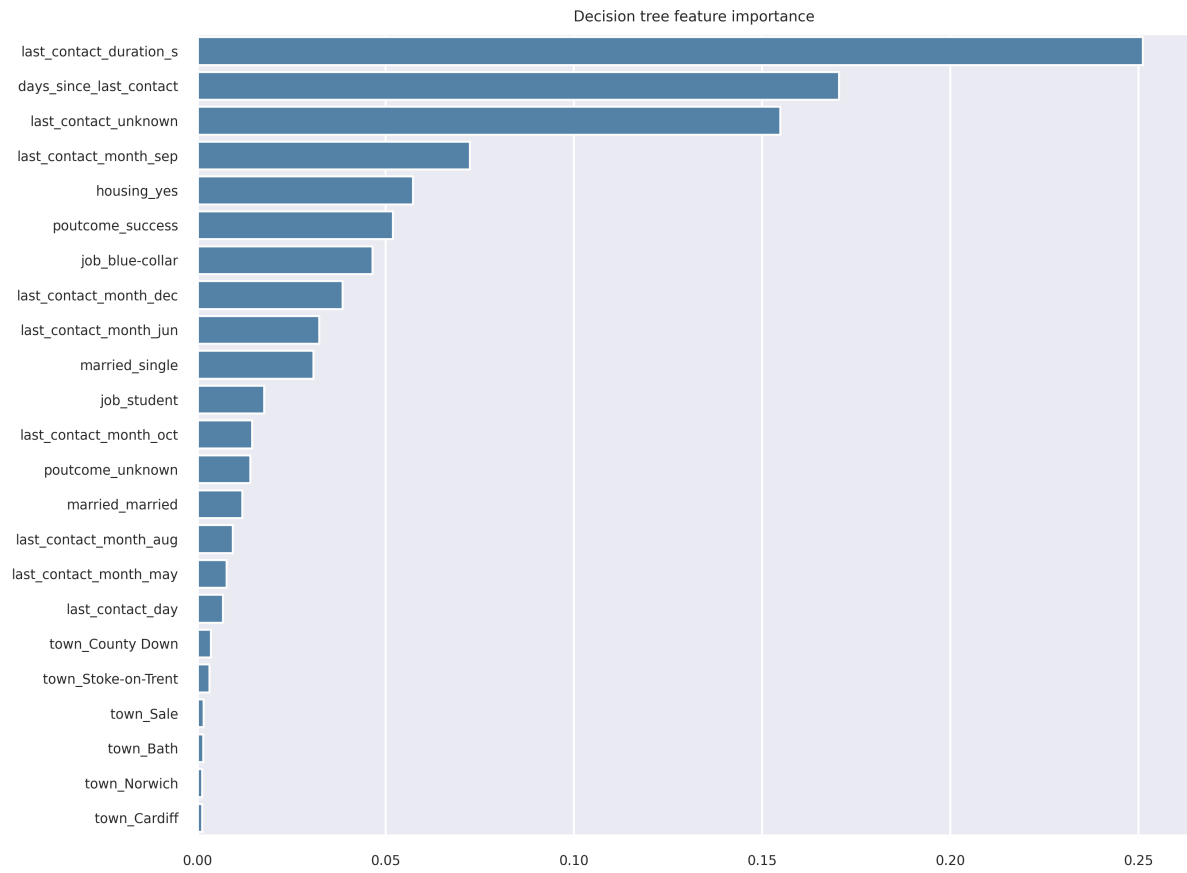   last_contact_month: True
        last_contact_duration_s: True
        campaign: True
        days_since_last_contact: True
        previous: True
        poutcome: True
        made_deposit: True

Decision tree feature importance

Dataset: bank-tr.csv
Experiment: exp_01

LogisticRegression(max_iter=1000, random_state=0)

LR experiment best results:

Experiment best score (accuracy): 0.820598

LR cross validation scores:
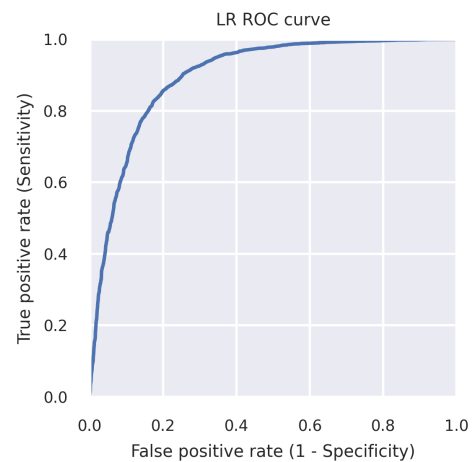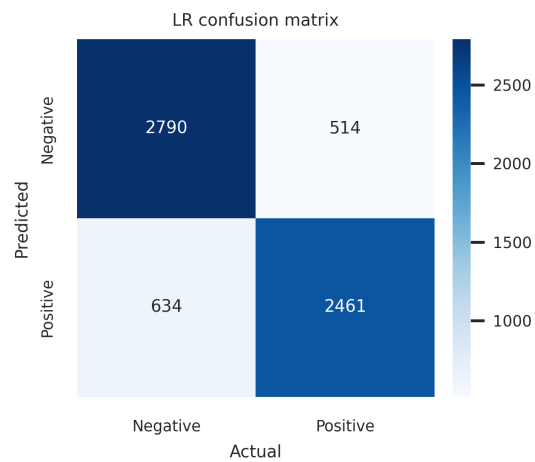
Accuracy: 0.820598
Precision: 0.827234
Recall: 0.795153
F1: 0.810819
AUC: 0.898059

LR confusion matrix & ROC curve:

Best classifier:

LogisticRegression(max_iter=1000, random_state=0)

Best hyperparameters:

C: 1.0
penalty: l1
solver: saga

Grid search hyperparameters:

penalty: ['l1', 'l2']
C: [1.0, 0.1, 10]
solver: ['liblinear', 'saga']

Dataset: bank-tr.csv

Experiment: exp_01

LogisticRegression(max_iter=1000, random_state=0)

Experiment parameters:
  n_splits: 5
  scoring: accuracy
  target: made_deposit
  categorical columns:
    town
    country
    job
    married
    education
    defaulted?
    housing
    has_loan
    last_contact
    cc_tr
    last_contact_month
    poutcome
  feature selection:
    accountID: False
    town: True
    country: True
    age: True
    job: True
    married: True
    education: True
    defaulted?: True
    current_balance: True
    housing: True
    has_loan: True
    last_contact: True
    cc_tr: True
    last_contact_day: True
    last_contact_month: True
    last_contact_duration_s: True
    campaign: True
    days_since_last_contact: True
    previous: True
    poutcome: True
    made_deposit: True

Logistic regression feature importance