Dataset: bank-tr.csv

Experiment: exp_01

DecisionTreeClassifier(random_state=0)

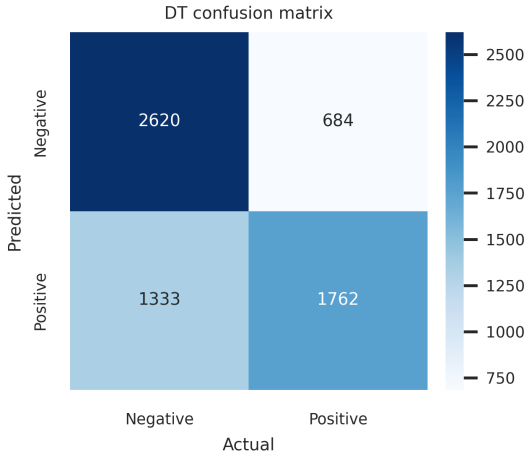
DT experiment best results:

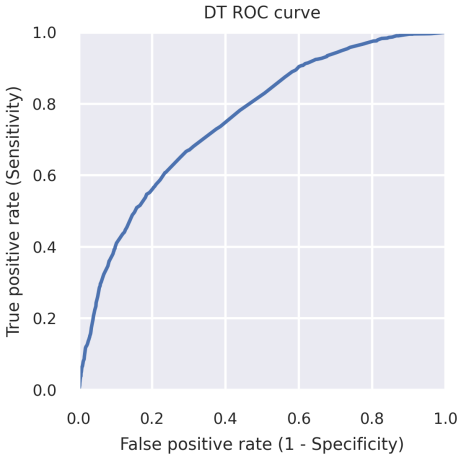
Experiment best score (accuracy): 0.684798

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

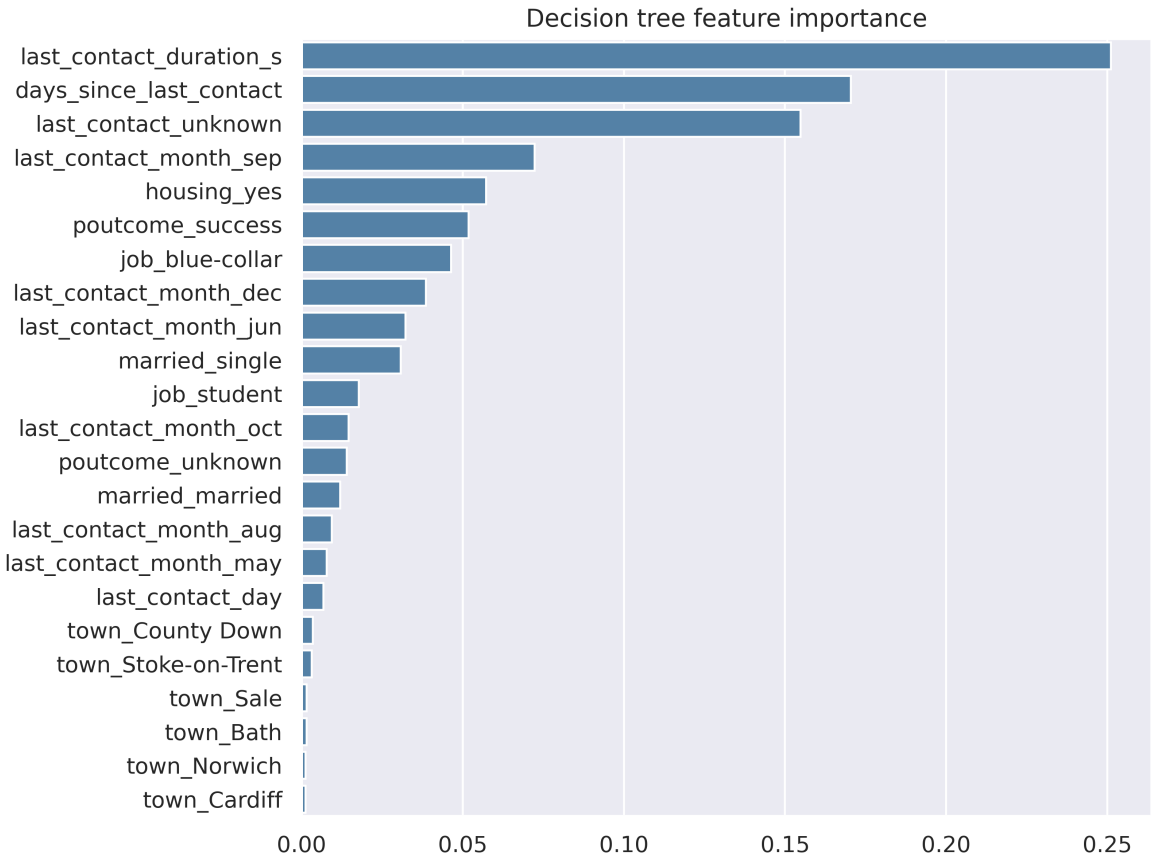
Experiment: exp_01

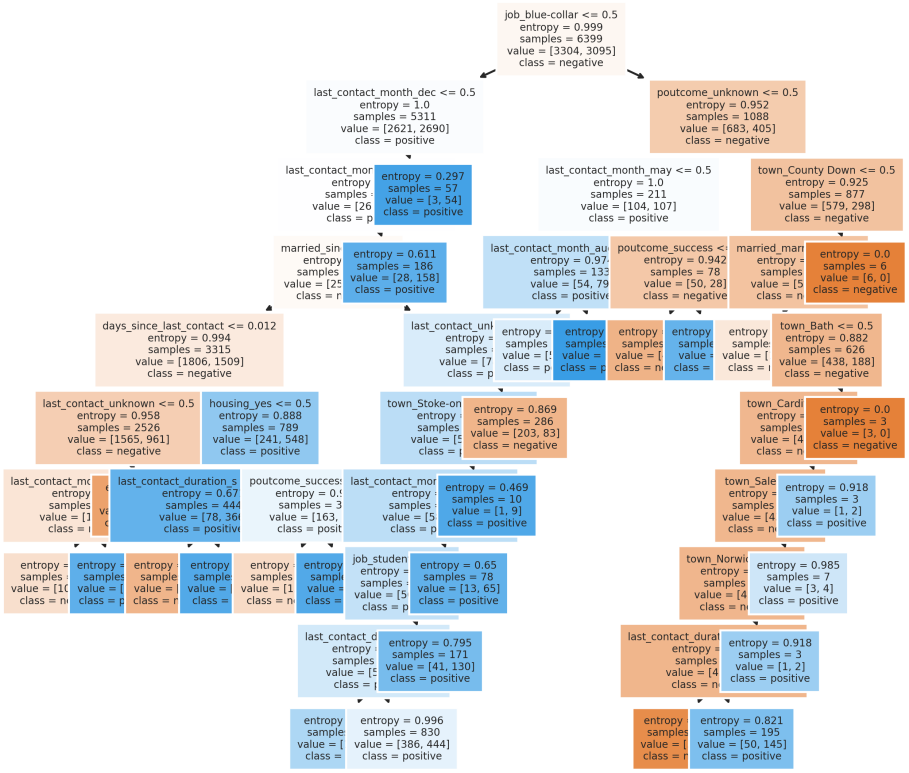
DecisionTreeClassifier(random_state=0)

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
last_contact_duration_s
days_since_last_contact
last_contact_unknown
last_contact_month_sep
housing_yes
poutcome_success
job_blue-collar
last_contact_month_dec
last_contact_month_jun
married_single
job_student
last_contact_month_oct
poutcome_unknown
married_married
last_contact_month_aug
last_contact_month_may
last_contact_day
town_County Down
town_Stoke-on-Trent
town_Sale
town_Bath
town_Norwich
town_Cardiff
0.00
0.05
0.10
0.15
0.20
0.25


job_blue-collar <= 0.5
entropy = 0.999
samples = 6399
value = [3304, 3095]
class = negative
last_contact_month_dec <= 0.5
entropy = 1.0
samples = 5311
value = [2621, 2690]
class = positive
poutcome_unknown <= 0.5
entropy = 0.952
samples = 1088
value = [683, 405]
class = negative
entropy = 0.297
samples = 57
value = [3, 54]
class = positive
last_contact_month_may <= 0.5
entropy = 1.0
samples = 211
value = [104, 107]
class = positive
town_County Down <= 0.5
entropy = 0.925
samples = 877
value = [579, 298]
class = negative
entropy = 0.611
samples = 186
value = [28, 158]
class = positive
entropy = 0.0
samples = 6
value = [6, 0]
class = negative
days_since_last_contact <= 0.012
entropy = 0.994
samples = 3315
value = [1806, 1509]
class = negative
town_Bath <= 0.5
entropy = 0.882
samples = 626
value = [438, 188]
class = negative
last_contact_unknown <= 0.5
entropy = 0.958
samples = 2526
value = [1565, 961]
class = negative
housing_yes <= 0.5
entropy = 0.888
samples = 789
value = [241, 548]
class = positive
entropy = 0.869
samples = 286
value = [203, 83]
class = negative
entropy = 0.0
samples = 3
value = [3, 0]
class = negative
entropy = 0.469
samples = 10
value = [1, 9]
class = positive
entropy = 0.918
samples = 3
value = [1, 2]
class = positive
entropy = 0.65
samples = 78
value = [13, 65]
class = positive
entropy = 0.985
samples = 7
value = [3, 4]
class = positive
entropy = 0.795
samples = 171
value = [41, 130]
class = positive
entropy = 0.918
samples = 3
value = [1, 2]
class = positive
entropy = 0.996
samples = 830
value = [386, 444]
class = positive
entropy = 0.821
samples = 195
value = [50, 145]
class = positive

Dataset: bank-tr.csv

Experiment: exp_01

LogisticRegression(max_iter=1000, random_state=0)

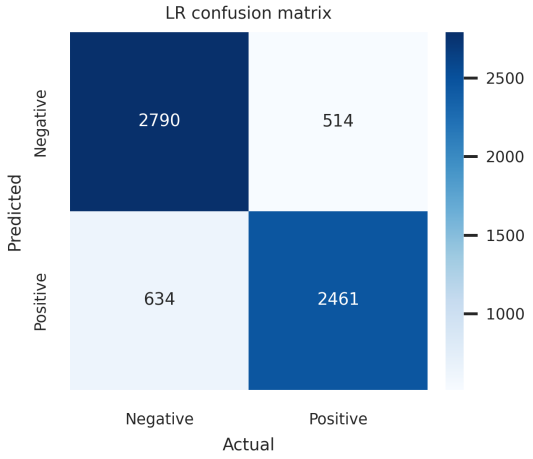
LR experiment best results:

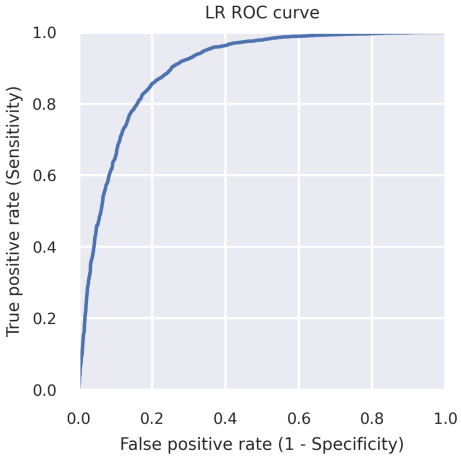
Experiment best score (accuracy): 0.820598

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

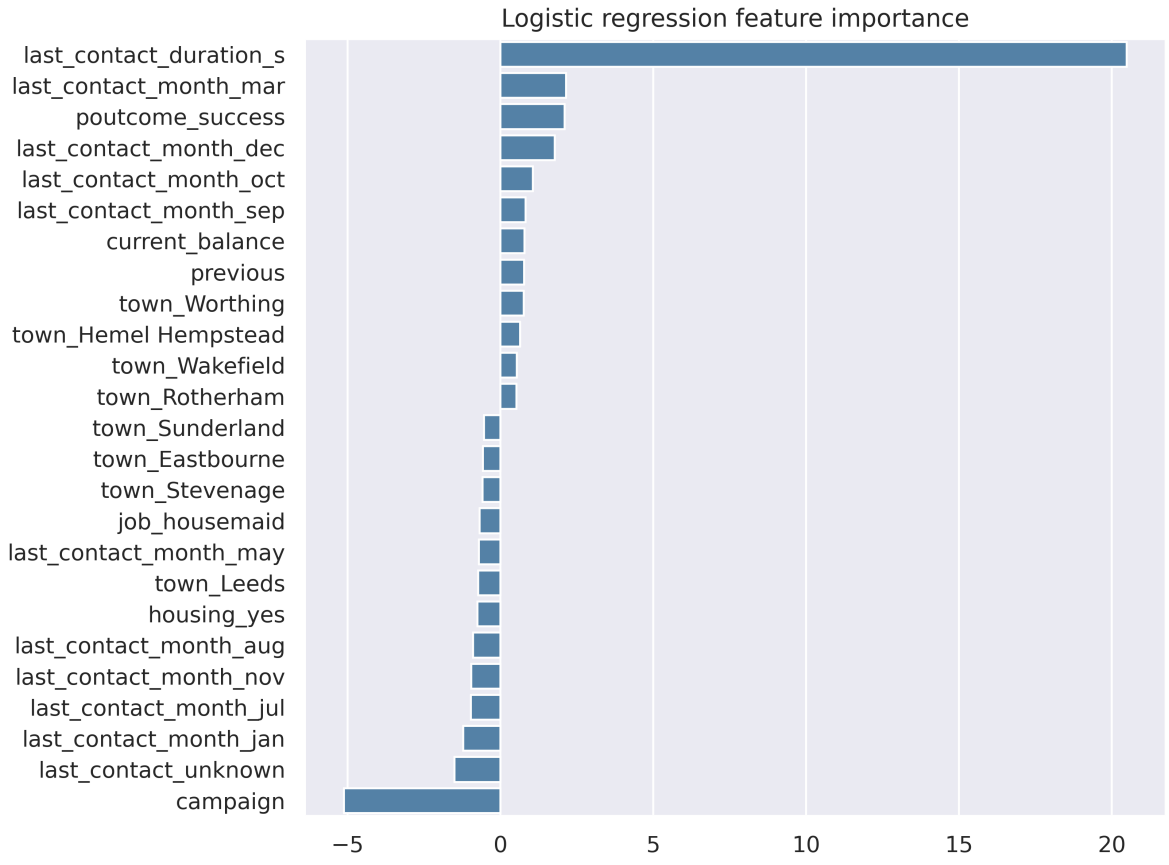
Experiment: exp_01

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
last_contact_duration_s
last_contact_month_mar
poutcome_success
last_contact_month_dec
last_contact_month_oct
last_contact_month_sep
current_balance
previous
town_Worthing
town_Hemel Hempstead
town_Wakefield
town_Rotherham
town_Sunderland
town_Eastbourne
town_Stevenage
job_housemaid
last_contact_month_may
town_Leeds
housing_yes
last_contact_month_aug
last_contact_month_nov
last_contact_month_jul
last_contact_month_jan
last_contact_unknown
campaign
−5
0
5
10
15
20