



ML training report for model: v01_CATBC_01_02

This report includes all the details on how the model was trained. This is the full version that also includes information on all the other models in the experiment.

Experiment

Name Churn Prediction	
Type	Binary
Target	Churn
Space	Personal
Creator	Rennie Crookston
Created	Jul 29, 2025 11:08 PM EDT
Versions	1
Models	9

Version information

Version 1 v1_20250730_2309	
Started	Jul 29, 2025 11:09 PM EDT
Finished	Jul 29, 2025 11:11 PM EDT
Algorithms	Elastic Net Regression, Gaussian Naive Bayes, LightGBM Classification, CatBoost Classification, Logistic Regression, Random Forest Classification, XGBoost Classification, Lasso Regression
Optimization	Intelligent
Time limit reached	False
Trained models	9
Batches	3

Model information

Name v01_CATBC_01_02	
Algorithm	CatBoost Classification
Created	Jul 29, 2025 11:10 PM EDT
Finished	Jul 29, 2025 11:11 PM EDT
Sampling ratio	100%

Training dataset

Name Telcom Churn Data Cleaned in SQL.csv	
Space	Default Data Space
Loaded	Jul 29, 2025 11:07 PM EDT
Rows	7,043
Columns	25
Cells	176,075

Table of contents

1. Model processing steps	3
2. Model feature treatment	4
3. Versions in experiment	5
4. Models in experiment	6
4.1 Model: v01_LGBMC_01_05	7
4.2 Model: v01_XGBC_01_04	9
4.3 Model: v01_RAFC_01_03	11
4.4 Model: v01_CATBC_01_02	12
4.5 Model: v01_LGBMC_01_01	14
4.6 Model: v01_LOGC_01_00	15
4.7 Model: v01_LOGC_00_02	16
4.8 Model: v01_LOGC_00_01	18
4.9 Model: v01_LGBMC_00_00	20
5. Model metrics	22
5.1 Metrics based on holdout rows	22
5.2 Metrics based on rows used for training	23
6. Training datasets	24
6.1 Dataset: Telcom Churn Data Cleaned in SQL.csv	24

1. Model processing steps

This table shows the steps taken to produce the final model: [v01_CATBC_01_02](#)

The table lists the changes introduced with each step, as well as the number of features included in the model after the step is completed.

Step	Changes	Features
Original dataset	<ul style="list-style-type: none">Telcom Churn Data Cleaned in SQL.csv	24
Feature engineering	<ul style="list-style-type: none">customerID was impact encodedgender, SeniorCitizen, Partner, Dependents, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, Churn, TenureMonthsBinned, HasPaperlessElectronicPayment were one-hot encoded	24
EDA	<ul style="list-style-type: none">customerID was dropped. Reason: High cardinalityFeature types of all features were identified, see Model feature treatment	23
Data preprocessing		23
Automated machine learning	<ul style="list-style-type: none">HasPaperlessElectronicPayment, gender, TenureMonthsBinned, SeniorCitizen, TotalServiceAddons, StreamingTV, Partner, Dependents, PhoneService, DeviceProtection, PaperlessBilling, PaymentMethod were dropped. Reason: Low permutation importance	10
Split	<ul style="list-style-type: none">A stratified subset (20%) was separated into a holdout	10
Training	<ul style="list-style-type: none">Five-fold cross-validationUsing the following algorithm: CatBoost ClassificationSetting hyperparameters	10

2. Model feature treatment

This table shows information and findings for each potential feature that could have been used for the model.

Potential feature	Included	Feature type (U=User override)	Encoding	Column characteristics
customerID	No	Categorical	Impact encoded	<ul style="list-style-type: none">High cardinality
gender	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
SeniorCitizen	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
Partner	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
Dependents	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
tenure	Yes	Numeric		
PhoneService	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
MultipleLines	Yes	Categorical	One-hot encoded	
InternetService	Yes	Categorical	One-hot encoded	
OnlineSecurity	Yes	Categorical	One-hot encoded	
OnlineBackup	Yes	Categorical	One-hot encoded	
DeviceProtection	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
TechSupport	Yes	Categorical	One-hot encoded	
StreamingTV	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
StreamingMovies	Yes	Categorical	One-hot encoded	
Contract	Yes	Categorical	One-hot encoded	
PaperlessBilling	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
PaymentMethod	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
MonthlyCharges	No	Numeric		<ul style="list-style-type: none">High correlation
TotalCharges	Yes	Numeric		
Churn	Yes	Categorical		
TenureMonthsBinned	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
TotalServiceAddons	No	Categorical		<ul style="list-style-type: none">Low permutation importance
HasPaperlessElectronicPayment	No	Categorical	One-hot encoded	<ul style="list-style-type: none">Low permutation importance
AverageMonthlyChargePerTenure	Yes	Numeric		



3. Versions in experiment

Version for selected model

Version	Name	Optimization	Models	Based on
---------	------	--------------	--------	----------

v1	v1_20250730_2309	Intelligent	9	Initial version
----	------------------	-------------	---	-----------------

Source dataset: [Telcom Churn Data Cleaned in SQL.csv](#)

4. Models in experiment

Selected model

Version	Name	Algorithm	F1	Optimization	Batch	Sampling	Features
v1	v01_CATBC_01_02	CatBoost Classification	0.613	Intelligent	2	100%	10

Other models

Version	Name	Algorithm	F1	Optimization	Batch	Sampling	Features
v1	v01_LGBMC_01_05	LightGBM Classification	0.610	Intelligent	2	100%	24
v1	v01_XGBC_01_04	XGBoost Classification	0.578	Intelligent	2	100%	10
v1	v01_RAFC_01_03	Random Forest Classification	0.603	Intelligent	2	100%	10
v1	v01_LGBMC_01_01	LightGBM Classification	0.601	Intelligent	2	100%	10
v1	v01_LOGC_01_00	Logistic Regression	0.609	Intelligent	2	100%	8
v1	v01_LOGC_00_02	Logistic Regression	0.609	Intelligent	1	100%	23
v1	v01_LOGC_00_01	Logistic Regression	0.607	Intelligent	1	100%	23
v1	v01_LGBMC_00_00	LightGBM Classification	0.614	Intelligent	1	100%	23

4.1 Model: v01_LGBMC_01_05

Algorithm	LightGBM Classification
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
TechSupport	Categorical	0.01553	0 (other)	One-hot encoded		0.00000
TotalCharges	Numeric	0.05427	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
MultipleLines	Categorical	0.00239	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric		0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
tenure	Numeric	0.29580	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
PaperlessBilling	Categorical	0.00383	0 (other)	One-hot encoded		0.00000
Partner	Categorical	0.00050	0 (other)	One-hot encoded		0.00000
MonthlyCharges	Numeric	0.03010	0 (mean)	Standardized	• 0.99624 - AverageMonthlyChargePerTenure	0.04347
StreamingTV	Categorical		0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.47900	0 (other)	One-hot encoded		0.00000
SeniorCitizen	Categorical	0.00016	0 (other)	One-hot encoded		0.00000
InternetService	Categorical	0.05331	0 (other)	One-hot encoded		0.00000
DeviceProtection	Categorical	0.00040	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical	0.00307	0 (other)	One-hot encoded		0.00000
Dependents	Categorical		0 (other)	One-hot encoded		0.00000
OnlineSecurity	Categorical	0.02834	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.00680	0 (other)	One-hot encoded		0.00000
TenureMonthsBinned	Categorical	0.00736	0 (other)	One-hot		0.00000



Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
				encoded		
PaymentMethod	Categorical	0.00584	0 (other)	One-hot encoded		0.00000
HasPaperlessElectronicPayment	Categorical	0.01301	0 (other)	One-hot encoded		0.00000
gender	Categorical		0 (other)	One-hot encoded		0.00000
PhoneService	Categorical	0.00029	0 (other)	One-hot encoded		0.00000
TotalServiceAddons	Categorical		0 (mean)	Standardized		0.03205

Hyperparameters

boosting_type	gbdt	class_weight	null
colsample_bytree	1.000	deterministic	0.000
force_col_wise	0.000	importance_type	split
learning_rate	0.100	max_depth	-1.000
min_child_samples	20.000	min_child_weight	0.001
min_split_gain	0.000	n_estimators	100.000
n_jobs	6.000	num_leaves	31.000
objective	null	random_state	42.000
reg_alpha	0.000	reg_lambda	0.000
subsample	1.000	subsample_for_bin	200,000.000
subsample_freq	0.000	verbosity	-1.000

4.2 Model: v01_XGBC_01_04

Algorithm	XGBoost Classification
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
AverageMonthlyChargePerTenure	Numeric	0.01543	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
Contract	Categorical	0.46202	0 (other)	One-hot encoded		0.00000
tenure	Numeric	0.27853	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
TechSupport	Categorical	0.02905	0 (other)	One-hot encoded		0.00000
TotalCharges	Numeric	0.10708	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
OnlineSecurity	Categorical	0.03677	0 (other)	One-hot encoded		0.00000
MultipleLines	Categorical	0.01943	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.00175	0 (other)	One-hot encoded		0.00000
InternetService	Categorical	0.04995	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical		0 (other)	One-hot encoded		0.00000

Hyperparameters

base_score	0.500	booster	gbtree
callbacks	null	colsample_bylevel	1.000
colsample_bynode	1.000	colsample_bytree	1.000
early_stopping_rounds	null	enable_categorical	0.000
eval_metric	null	gamma	0.000
gpu_id	-1.000	grow_policy	depthwise
importance_type	null	interaction_constraints	0.000
learning_rate	0.300	max_bin	256.000
max_cat_to_onehot	4.000	max_delta_step	0.000
max_depth	6.000	max_leaves	0.000
min_child_weight	1.000	missing	null
monotone_constraints	()	n_estimators	100.000



n_jobs	6.000	num_parallel_tree	1.000
objective	binary:logistic	predictor	auto
random_state	42.000	reg_alpha	0.000
reg_lambda	1.000	sampling_method	uniform
scale_pos_weight	1.000	subsample	1.000
tree_method	exact	use_label_encoder	0.000
validate_parameters	1.000	verbosity	null

4.3 Model: v01_RAFC_01_03

Algorithm	Random Forest Classification
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
TotalCharges	Numeric	0.12769	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
MultipleLines	Categorical	0.00598	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical		0 (other)	One-hot encoded		0.00000
InternetService	Categorical	0.12999	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical		0 (other)	One-hot encoded		0.00000
tenure	Numeric	0.15531	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
OnlineSecurity	Categorical	0.04060	0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.47211	0 (other)	One-hot encoded		0.00000
TechSupport	Categorical	0.05873	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric	0.00959	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372

Hyperparameters

bootstrap	0.000	ccp_alpha	0.000
class_weight	null	criterion	gini
max_depth	10.000	max_features	sqrt
max_leaf_nodes	null	max_samples	null
min_impurity_decrease	0.000	min_samples_leaf	1.000
min_samples_split	2.000	min_weight_fraction_leaf	0.000
monotonic_cst	null	n_estimators	100.000
n_jobs	6.000	oob_score	0.000
random_state	42.000	verbose	0.000
warm_start	0.000		

4.4 Model: v01_CATBC_01_02

Algorithm	CatBoost Classification
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
tenure	Numeric	0.32910	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
MultipleLines	Categorical	0.03152	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric	0.02896	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
TotalCharges	Numeric	0.05070	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
InternetService	Categorical	0.06236	0 (other)	One-hot encoded		0.00000
OnlineSecurity	Categorical	0.04023	0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.41245	0 (other)	One-hot encoded		0.00000
TechSupport	Categorical	0.02985	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical		0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.01483	0 (other)	One-hot encoded		0.00000

Hyperparameters

auto_class_weights	None	bayesian_matrix_reg	0.100
best_model_min_trees	1.000	boost_from_average	0.000
boosting_type	Plain	bootstrap_type	MVS
border_count	254.000	class_names	01
classes_count	0.000	depth	6.000
eval_fraction	0.000	eval_metric	Logloss
feature_border_type	GreedyLogSum	force_unit_auto_pair_weights	0.000
grow_policy	SymmetricTree	iterations	1,000.000
l2_leaf_reg	3.000	leaf_estimation_backtracking	AnyImprovement
leaf_estimation_iterations	10.000	leaf_estimation_method	Newton
learning_rate	0.022	loss_function	Logloss
max_leaves	64.000	min_data_in_leaf	1.000



model_shrink_mode	Constant	model_shrink_rate	0.000
model_size_reg	0.500	nan_mode	Min
penalties_coefficient	1.000	posterior_sampling	0.000
random_seed	42.000	random_strength	1.000
rsm	1.000	sampling_frequency	PerTree
score_function	Cosine	sparse_features_conflict_fraction	0.000
subsample	0.800	task_type	CPU
use_best_model	0.000		

4.5 Model: v01_LGBMC_01_01

Algorithm	LightGBM Classification
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
InternetService	Categorical	0.05655	0 (other)	One-hot encoded		0.00000
tenure	Numeric	0.29955	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
OnlineSecurity	Categorical	0.03227	0 (other)	One-hot encoded		0.00000
TechSupport	Categorical	0.02627	0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.42969	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.00838	0 (other)	One-hot encoded		0.00000
TotalCharges	Numeric	0.04708	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
StreamingMovies	Categorical	0.00080	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric	0.06736	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
MultipleLines	Categorical	0.03205	0 (other)	One-hot encoded		0.00000

Hyperparameters

boosting_type	gbdt	class_weight	null
colsample_bytree	1.000	deterministic	0.000
force_col_wise	0.000	importance_type	split
learning_rate	0.100	max_depth	-1.000
min_child_samples	20.000	min_child_weight	0.001
min_split_gain	0.000	n_estimators	100.000
n_jobs	6.000	num_leaves	31.000
objective	null	random_state	42.000
reg_alpha	0.000	reg_lambda	0.000
subsample	1.000	subsample_for_bin	200,000.000
subsample_freq	0.000	verbosity	-1.000

4.6 Model: v01_LOGC_01_00

Algorithm	Logistic Regression
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
Contract	Categorical	0.11461	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric	0.01883	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
InternetService	Categorical	0.01326	0 (other)	One-hot encoded		0.00000
tenure	Numeric	0.72168	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
TotalCharges	Numeric	0.05586	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
OnlineSecurity	Categorical	0.02591	0 (other)	One-hot encoded		0.00000
TenureMonthsBinned	Categorical	0.03413	0 (other)	One-hot encoded		0.00000
TechSupport	Categorical	0.01573	0 (other)	One-hot encoded		0.00000

Hyperparameters

C	1.000	class_weight	null
dual	0.000	fit_intercept	0.000
intercept_scaling	1.000	l1_ratio	null
max_iter	100.000	multi_class	deprecated
n_jobs	null	penalty	l2
random_state	42.000	solver	lbfgs
tol	0.000	verbose	0.000
warm_start	0.000		

4.7 Model: v01_LOGC_00_02

Algorithm	Logistic Regression
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
Dependents	Categorical	0.00135	0 (other)	One-hot encoded		0.00000
TotalCharges	Numeric	0.04061	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
OnlineSecurity	Categorical	0.01372	0 (other)	One-hot encoded		0.00000
MultipleLines	Categorical	0.00492	0 (other)	One-hot encoded		0.00000
tenure	Numeric	0.73951	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
gender	Categorical		0 (other)	One-hot encoded		0.00000
Partner	Categorical	0.00006	0 (other)	One-hot encoded		0.00000
PhoneService	Categorical	0.00275	0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.09467	0 (other)	One-hot encoded		0.00000
SeniorCitizen	Categorical	0.00219	0 (other)	One-hot encoded		0.00000
DeviceProtection	Categorical	0.00075	0 (other)	One-hot encoded		0.00000
TechSupport	Categorical	0.00883	0 (other)	One-hot encoded		0.00000
StreamingTV	Categorical		0 (other)	One-hot encoded		0.00000
PaymentMethod	Categorical	0.00121	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric	0.00739	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
InternetService	Categorical	0.03478	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical		0 (other)	One-hot encoded		0.00000
PaperlessBilling	Categorical	0.00325	0 (other)	One-hot encoded		0.00000



Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
TotalServiceAddons	Categorical		0 (mean)	Standardized		0.03205
HasPaperlessElectronicPayment	Categorical	0.00265	0 (other)	One-hot encoded		0.00000
TenureMonthsBinned	Categorical	0.03730	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.00406	0 (other)	One-hot encoded		0.00000

Hyperparameters

C	1.000	class_weight	null
dual	0.000	fit_intercept	0.000
intercept_scaling	1.000	l1_ratio	null
max_iter	100.000	multi_class	deprecated
n_jobs	null	penalty	l2
random_state	42.000	solver	lbfgs
tol	0.000	verbose	0.000
warm_start	0.000		

4.8 Model: v01_LOGC_00_01

Algorithm	Logistic Regression
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
Dependents	Categorical	0.00371	0 (other)	One-hot encoded		0.00000
TenureMonthsBinned	Categorical	0.05660	0 (other)	One-hot encoded		0.00000
Partner	Categorical		0 (other)	One-hot encoded		0.00000
OnlineSecurity	Categorical	0.04416	0 (other)	One-hot encoded		0.00000
MultipleLines	Categorical	0.01183	0 (other)	One-hot encoded		0.00000
DeviceProtection	Categorical	0.00332	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.01567	0 (other)	One-hot encoded		0.00000
TotalCharges	Numeric	0.04457	0 (mean)	Binning applied	• 0.82618 - tenure	0.04363
SeniorCitizen	Categorical	0.00680	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical		0 (other)	One-hot encoded		0.00000
StreamingTV	Categorical		0 (other)	One-hot encoded		0.00000
HasPaperlessElectronicPayment	Categorical	0.01025	0 (other)	One-hot encoded		0.00000
PhoneService	Categorical	0.00558	0 (other)	One-hot encoded		0.00000
PaperlessBilling	Categorical	0.00932	0 (other)	One-hot encoded		0.00000
TotalServiceAddons	Categorical		0 (mean)	Binning applied		0.03205
TechSupport	Categorical	0.02778	0 (other)	One-hot encoded		0.00000
PaymentMethod	Categorical	0.00305	0 (other)	One-hot encoded		0.00000



Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
tenure	Numeric	0.29434	0 (mean)	Binning applied	• 0.82618 - TotalCharges	0.07699
AverageMonthlyChargePerTenure	Numeric		0 (mean)	Binning applied	• 0.99624 - MonthlyCharges	0.03372
gender	Categorical		0 (other)	One-hot encoded		0.00000
InternetService	Categorical	0.10115	0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.36187	0 (other)	One-hot encoded		0.00000

Hyperparameters

C	1.000	class_weight	null
dual	0.000	fit_intercept	0.000
intercept_scaling	1.000	l1_ratio	null
max_iter	100.000	multi_class	deprecated
n_jobs	null	penalty	l2
random_state	42.000	solver	lbfgs
tol	0.000	verbose	0.000
warm_start	0.000		

4.9 Model: v01_LGBMC_00_00

Algorithm	LightGBM Classification
Version	1
Optimization	Intelligent
Sampling ratio	100%
Source dataset	Telcom Churn Data Cleaned in SQL.csv

This table includes the features used by the model

Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
Dependents	Categorical		0 (other)	One-hot encoded		0.00000
Contract	Categorical	0.48043	0 (other)	One-hot encoded		0.00000
tenure	Numeric	0.20655	0 (mean)	Standardized	• 0.82618 - TotalCharges	0.07699
InternetService	Categorical	0.06209	0 (other)	One-hot encoded		0.00000
PhoneService	Categorical		0 (other)	One-hot encoded		0.00000
TechSupport	Categorical	0.01943	0 (other)	One-hot encoded		0.00000
StreamingMovies	Categorical	0.01025	0 (other)	One-hot encoded		0.00000
TotalServiceAddons	Categorical	0.00079	0 (mean)	Standardized		0.03205
Partner	Categorical		0 (other)	One-hot encoded		0.00000
TotalCharges	Numeric	0.09539	0 (mean)	Standardized	• 0.82618 - tenure	0.04363
OnlineSecurity	Categorical	0.02940	0 (other)	One-hot encoded		0.00000
MultipleLines	Categorical	0.01368	0 (other)	One-hot encoded		0.00000
TenureMonthsBinned	Categorical	0.00425	0 (other)	One-hot encoded		0.00000
AverageMonthlyChargePerTenure	Numeric	0.04950	0 (mean)	Standardized	• 0.99624 - MonthlyCharges	0.03372
gender	Categorical	0.00645	0 (other)	One-hot encoded		0.00000
OnlineBackup	Categorical	0.00959	0 (other)	One-hot encoded		0.00000
PaymentMethod	Categorical		0 (other)	One-hot encoded		0.00000
SeniorCitizen	Categorical	0.00273	0 (other)	One-hot encoded		0.00000
StreamingTV	Categorical	0.00034	0 (other)	One-hot		0.00000



Feature	Feature type	Importance	Imputation	Encoding	Feature correlation	Target correlation
				encoded		
PaperlessBilling	Categorical		0 (other)	One-hot encoded		0.00000
HasPaperlessElectronicPayment	Categorical	0.00913	0 (other)	One-hot encoded		0.00000
DeviceProtection	Categorical		0 (other)	One-hot encoded		0.00000

Hyperparameters

boosting_type	gbdt	class_weight	null
colsample_bytree	1.000	deterministic	0.000
force_col_wise	0.000	importance_type	split
learning_rate	0.100	max_depth	-1.000
min_child_samples	20.000	min_child_weight	0.001
min_split_gain	0.000	n_estimators	100.000
n_jobs	6.000	num_leaves	31.000
objective	null	random_state	42.000
reg_alpha	0.000	reg_lambda	0.000
subsample	1.000	subsample_for_bin	200,000.000
subsample_freq	0.000	verbosity	-1.000

5. Model metrics

5.1 Metrics based on holdout rows

Model	F1	AUC	Precision	Accuracy	Recall	Prediction speed ¹	Miss Rate	Fallout	Specificity	NPV	MCC	Threshold	Log loss
v01_LGBMC_01_05	0.61003	0.82416	0.49420	0.72960	0.79679	30339	0.20321	0.29469	0.70531	0.90571	0.44810	0.23645	0.44766
v01_XGBC_01_04	0.57751	0.79948	0.46493	0.70405	0.76203	63613	0.23797	0.31691	0.68309	0.88819	0.39646	0.20646	0.48318
v01_RAFC_01_03	0.60338	0.82304	0.49826	0.73314	0.76471	33303	0.23529	0.27826	0.72174	0.89461	0.43716	0.28364	0.44797
v01_CATBC_01_02	0.61336	0.82677	0.51948	0.74947	0.74866	71323	0.25134	0.25024	0.74976	0.89195	0.45284	0.28866	0.44165
v01_LGBMC_01_01	0.60080	0.82621	0.47930	0.71611	0.80481	67068	0.19519	0.31594	0.68406	0.90653	0.43430	0.23030	0.44315
v01_LOGC_01_00	0.60889	0.82250	0.53082	0.75656	0.71390	26614	0.28610	0.22802	0.77198	0.88190	0.44781	0.33764	0.44193
v01_LOGC_00_02	0.60948	0.83018	0.52734	0.75444	0.72193	24923	0.27807	0.23382	0.76618	0.88406	0.44812	0.30402	0.43380
v01_LOGC_00_01	0.60730	0.83518	0.50717	0.74024	0.75668	21673	0.24332	0.26570	0.73430	0.89307	0.44329	0.26355	0.42874
v01_LGBMC_00_00	0.61392	0.82524	0.50697	0.74024	0.77807	31826	0.22193	0.27343	0.72657	0.90060	0.45352	0.24910	0.44685

1. Prediction speed refers to how fast a machine learning model is able to generate predictions. In Qlik Predict, prediction speed is calculated using the combined feature computing time and test dataset prediction time. It is displayed in rows per second.



5.2 Metrics based on rows used for training

Model	F1	AUC	Precision	Accuracy	Recall	Miss Rate	Fallout	Specificity	NPV	MCC	Threshold	Log loss
v01_LGBMC_01_05	0.61987	0.83444	0.51316	0.74530	0.78261	0.21739	0.26818	0.73182	0.90310	0.46275	0.23645	0.43760
v01_XGBC_01_04	0.60971	0.82025	0.49643	0.73163	0.78997	0.21003	0.28944	0.71056	0.90353	0.44743	0.20646	0.46775
v01_RAFC_01_03	0.61592	0.83287	0.51386	0.74565	0.76856	0.23144	0.26262	0.73738	0.89818	0.45658	0.28364	0.43449
v01_CATBC_01_02	0.62990	0.83973	0.53590	0.76180	0.76388	0.23612	0.23895	0.76105	0.89923	0.47793	0.28866	0.42393
v01_LGBMC_01_01	0.61935	0.83036	0.50420	0.73820	0.80268	0.19732	0.28509	0.71491	0.90934	0.46265	0.23030	0.44195
v01_LOGC_01_00	0.63313	0.84222	0.55040	0.77086	0.74515	0.25485	0.21986	0.78014	0.89446	0.48340	0.33764	0.42028
v01_LOGC_00_02	0.64000	0.84871	0.54727	0.76997	0.77057	0.22943	0.23025	0.76975	0.90281	0.49314	0.30402	0.41225
v01_LOGC_00_01	0.63780	0.84936	0.53117	0.75950	0.79799	0.20201	0.25441	0.74559	0.91086	0.49018	0.26355	0.41170
v01_LGBMC_00_00	0.62050	0.83488	0.51815	0.74902	0.77324	0.22676	0.25972	0.74028	0.90038	0.46360	0.24910	0.43701

6. Training datasets

6.1 Dataset: Telcom Churn Data Cleaned in SQL.csv

This dataset is used in versions: 1

The following table contains information about the source dataset and how it is interpreted by the experiment. It also includes automatically engineered features.

Column	Data type	Distinct values	Null values	Sample values
customerID	String	7043	0	0002-ORFBO (1), 0003-MKNFE (1), 0004-TLHLJ (1), 0011-IGKFF (1), 0013-EXCHZ (1)
gender	String	2	0	Male (3555), Female (3488)
SeniorCitizen	String	2	0	No (5901), Yes (1 142)
Partner	String	2	0	No (3641), Yes (3402)
Dependents	String	2	0	No (4933), Yes (2110)
tenure	Integer	73	0	1 (613), 72 (362), 2 (238), 3 (200), 4 (176)
PhoneService	String	2	0	Yes (6361), No (682)
MultipleLines	String	3	0	No (3390), Yes (2971), No phone service (682)
InternetService	String	3	0	Fiber optic (3096), DSL (2421), No (1526)
OnlineSecurity	String	3	0	No (3498), Yes (2019), No internet service (1526)
OnlineBackup	String	3	0	No (3088), Yes (2429), No internet service (1526)
DeviceProtection	String	3	0	No (3095), Yes (2422), No internet service (1526)
TechSupport	String	3	0	No (3473), Yes (2044), No internet service (1526)
StreamingTV	String	3	0	No (2810), Yes (2707), No internet service (1526)
StreamingMovies	String	3	0	No (2785), Yes (2732), No internet service (1526)
Contract	String	3	0	Month-to-month (3875), Two year (1695), One year (1473)
PaperlessBilling	String	2	0	Yes (4171), No (2872)
PaymentMethod	String	4	0	Electronic check (2365), Mailed check (1612), Bank transfer (automatic) (1544), Credit card (automatic) (1522)
MonthlyCharges	Float (Double)	1585	0	20.05 (61), 19.85 (45), 19.95 (44), 19.9 (44), 19.7 (43)
TotalCharges	Float (Double)	6531	0	0 (11), 20.2 (11), 19.75 (9), 19.65 (8), 20.05 (8)
Churn	String	2	0	No (5174), Yes (1869)
TenureMonthsBinned	String	4	0	49-72 Months (2239), 0-12 Months (2186), 25-48 Months (1594), 13-24 Months (1024)
TotalServiceAddons	Integer	7	0	0 (2219), 3 (1118), 2 (1033), 1 (966), 4 (852)
HasPaperlessElectronicPayment	String	2	0	No (5301), Yes (1742)
AverageMonthlyChargePerTenure	Float (Double)	6546	0	20.2 (12), 19.65 (11), 19.9 (10), 19.75 (9), 20.45 (8)