National Research Tomsk State University

Лаборатория анализа данных физики высоких энергий Томского государственного университета

Measurement of differential cross-sections of a single top quark produced in association with a W boson with ATLAS at $\sqrt{s} = 13$ TeV

Progress Report

Neda Firoz

Goal: separate **tW (top+anti-top)** from **$t\bar{t}$** in the **1j1b dilepton** region

•**Inputs (9 variables used):**
bdt_centrality_ll_recalc_NOSYS,
bdt_delta_pT_ll_MET_recalc_NOSYS,
bdt_delta_pT_llb_MET_recalc_NOSYS,
bdt_eta_llMetB_recalc_NOSYS,
bdt_m_l1b_recalc_NOSYS,
bdt_m_l2b_recalc_NOSYS,
bdt_pT_llMetB_recalc_NOSYS,
bdt_pT_llb_recalc_NOSYS,
bdt_sum_ET_recalc_NOSYS.

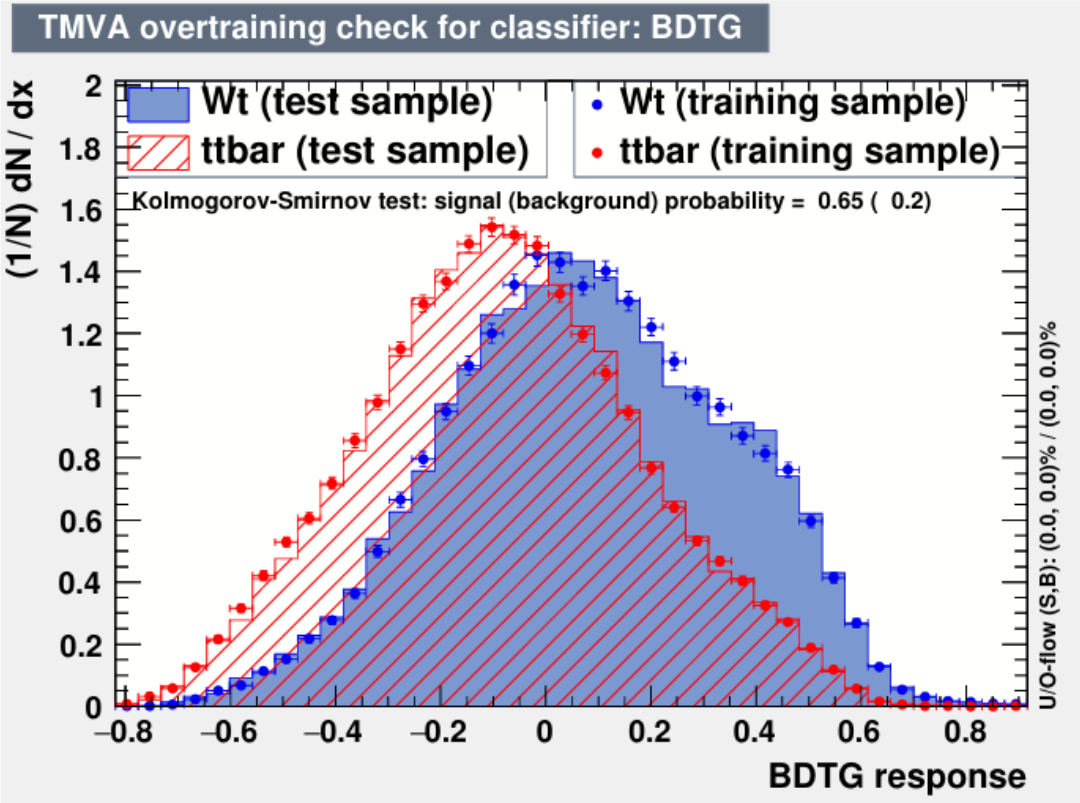
•**Samples / tree:** all files' tree name is analysis

Signal: tW (top) + $\bar{t}$W (anti-top)
Background: $t\bar{t}$ (non-all-had)

•**Event weights:** auto-resolved to weight_mc_NOSYS * weight_pileup_NOSYS * globalTriggerEffSF_NOSYS.

•**Bad values:** any variable ≤ **−990** or non-finite is masked per event.

Article's Report on BDT



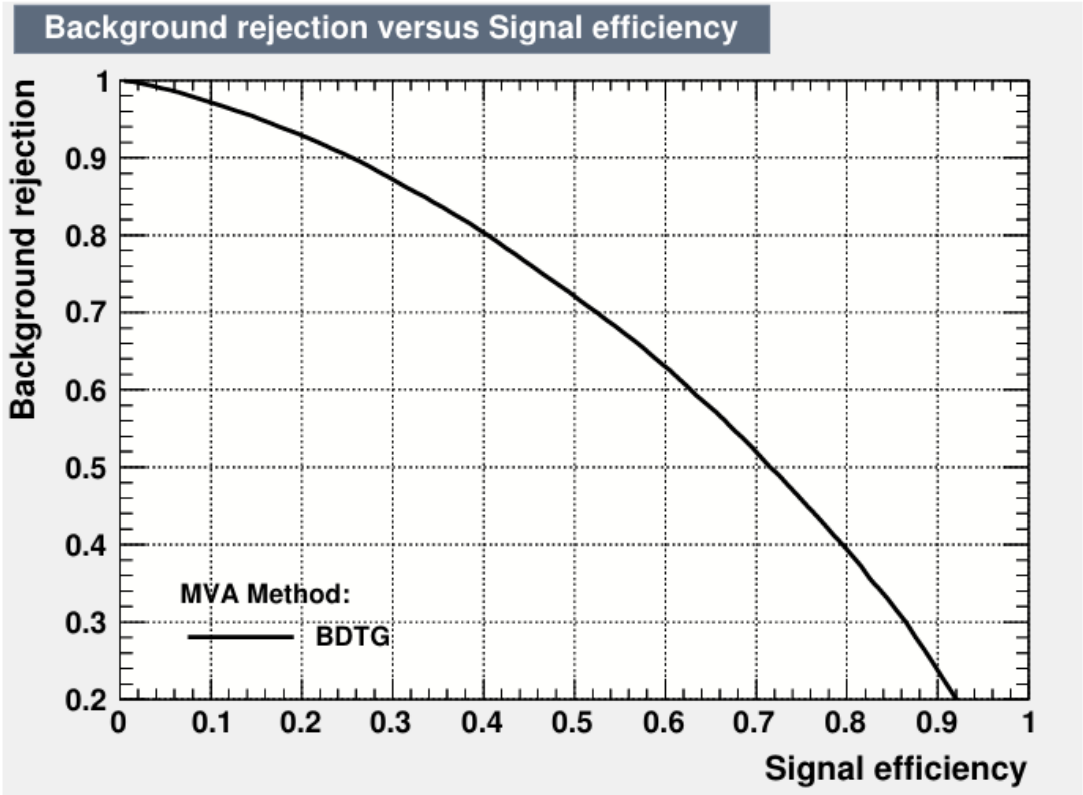


Figure 9: Comparison of test/training sample distributions and background rejection factor versus signal efficiency.

RESULTS on NEW DATA after some hyperparameters optimization and k-fold 20.01.2026

K-fold Validation (TMVA)

Classes / raw events added

Signal: 48,776

Background: 266,006

After != -999 preselection, TMVA reports:

Signal passed: 12,492

Background passed: 75,074

Train/Test trees :

TrainTree = 61,297,

TestTree = 26,269

Model comparison (ROC integral on test sample)

Ranked by TMVA

1.BDTG: 0.665

2.ANN: 0.661

3.BDT: 0.660

4.KNN: 0.648

5.Fisher: 0.632

6.Likelihood: 0.629

Best AUC/ROC integral ≈ 0.665 with BDTG

Preliminary result set-TMVA

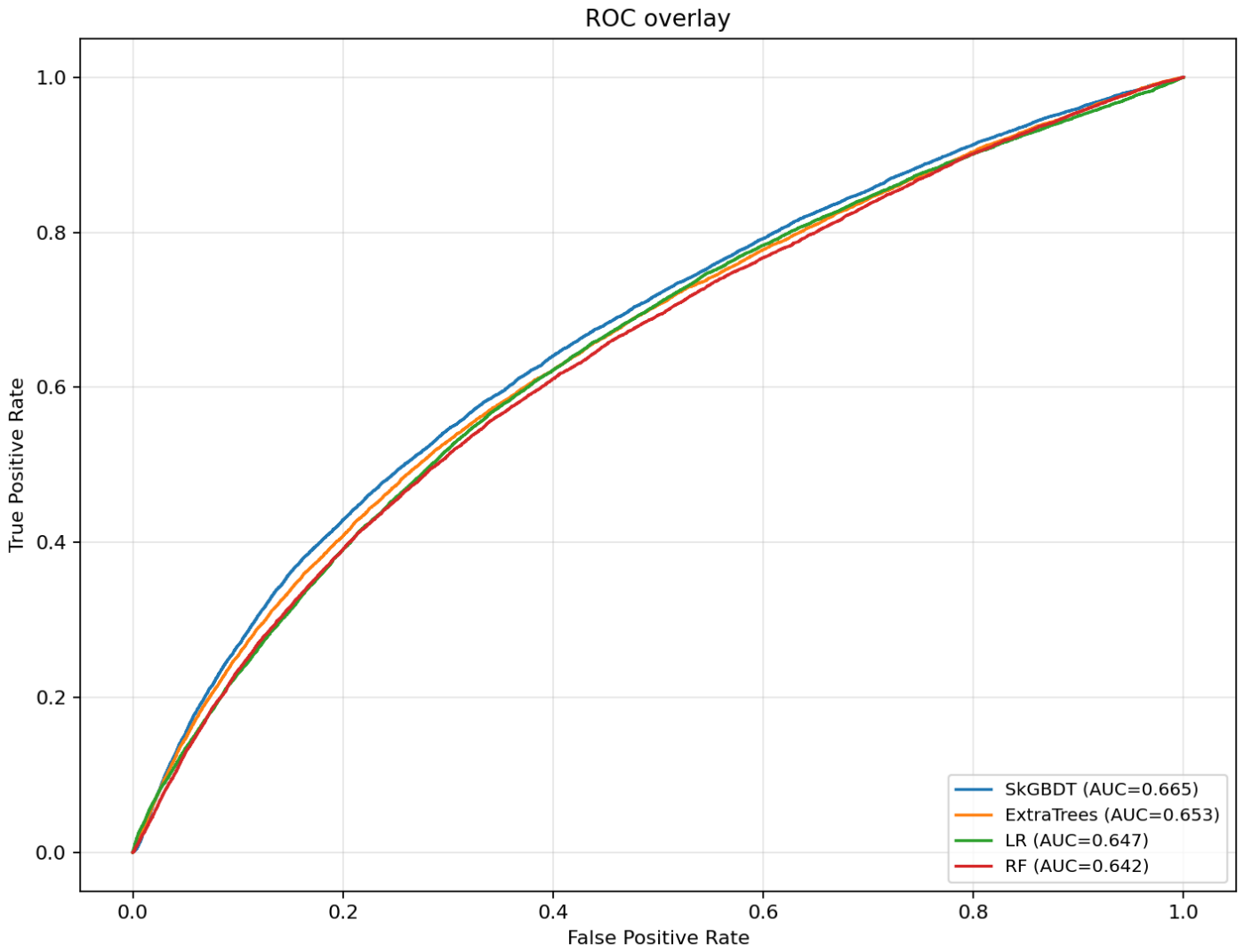
Method	B = 0.01	B = 0.10	B = 0.30
BDTG	0.041 (**0.047**)	0.272 (**0.268**)	0.541 (**0.541**)
ANN	0.041 (**0.048**)	0.272 (**0.267**)	0.539 (**0.535**)
BDT	0.045 (**0.049**)	0.261 (**0.262**)	0.538 (**0.539**)
KNN	0.038 (**0.042**)	0.249 (**0.254**)	0.527 (**0.528**)
Fisher	0.033 (**0.034**)	0.214 (**0.202**)	0.482 (**0.478**)
Likelihood	0.022 (**0.023**)	0.208 (**0.207**)	0.489 (**0.489**)

- Table shows the signal efficiency at fixed background efficiencies of 1%, 10%, and 30%.
- Boosted decision trees (BDTG and BDT) consistently achieve the highest signal efficiency across all operating points.
- The agreement between training and test efficiencies indicates no significant overtraining, particularly for BDTG and BDT.”

Preliminary result set for K-fold Validation in Python

Model	AUC_mean	AUC_std	kfold	n_events	n_valid_folds
SketchboostGBDT	0.665143	0.005386	5	87566	5
ExtraTrees	0.652857	0.004632	5	87566	5
LR	0.647498	0.004209	5	87566	5
RF	0.641839	0.003821	5	87566	5

K-fold on Python Trial 1

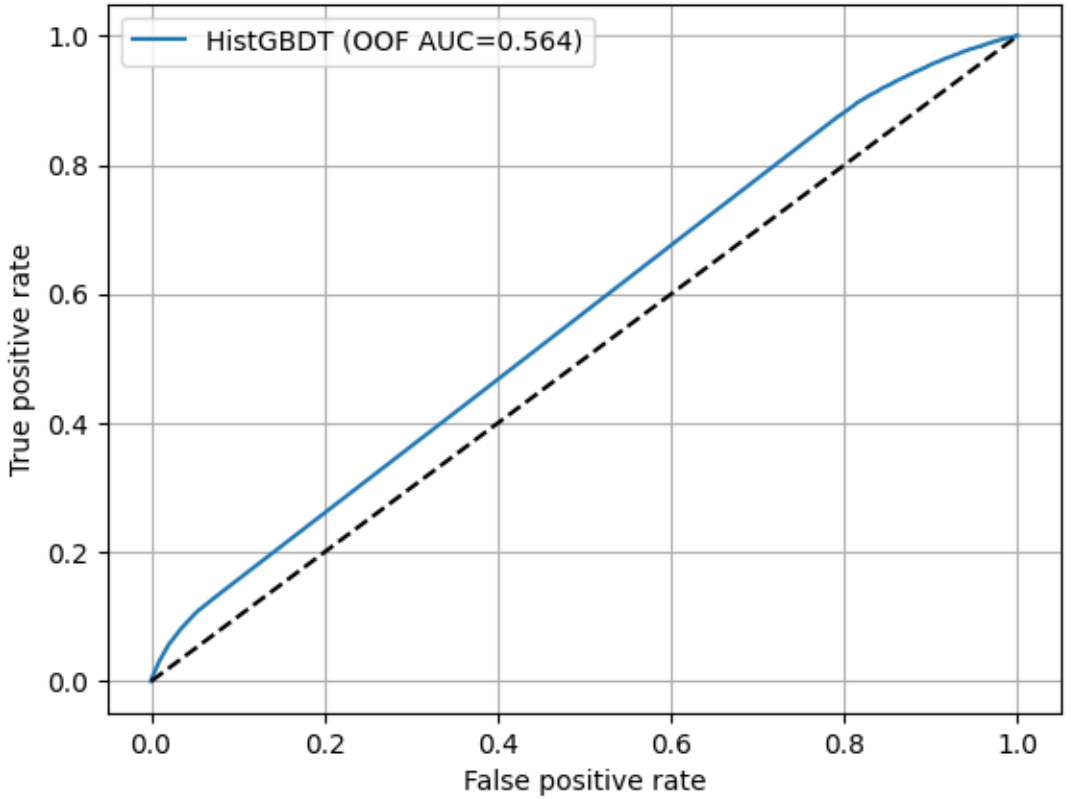


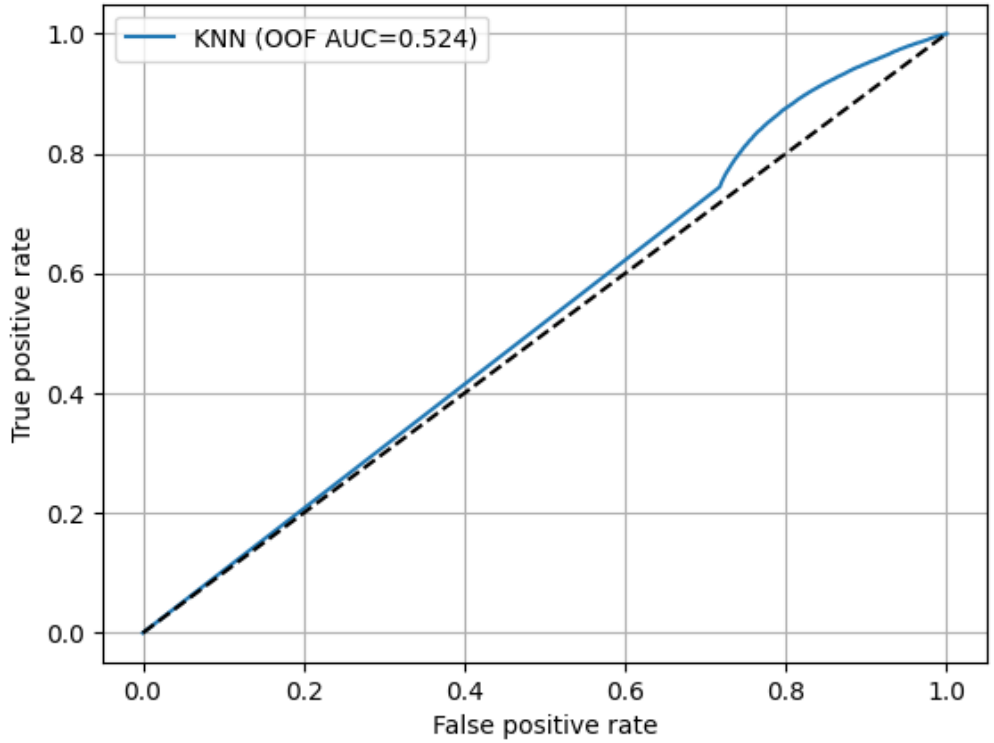
Performance of ML Algorithms using K-fold on Python Trial 2

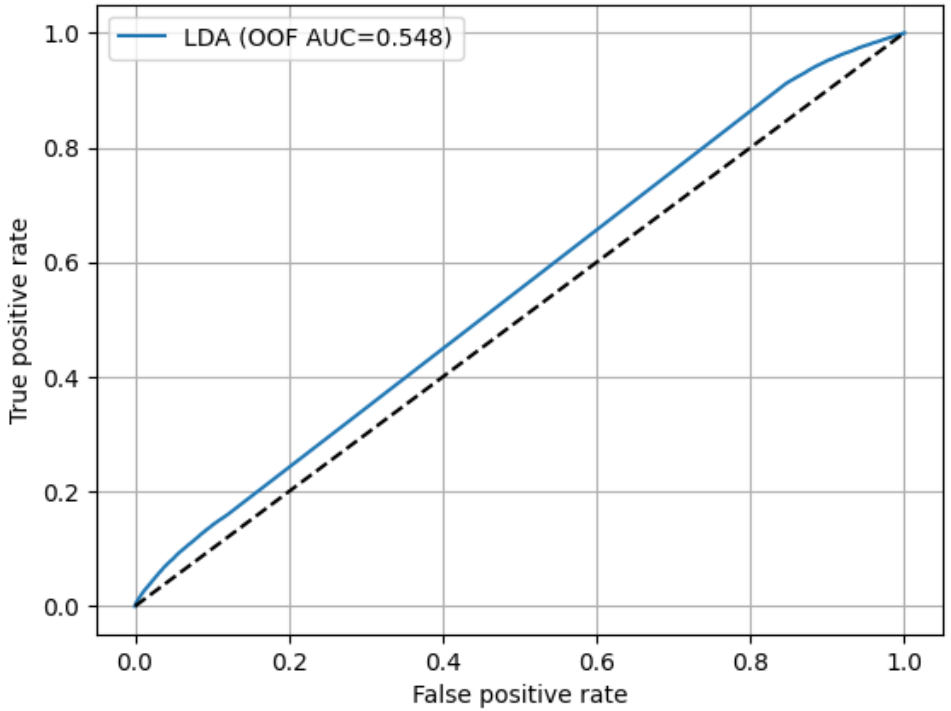
20.01.2026

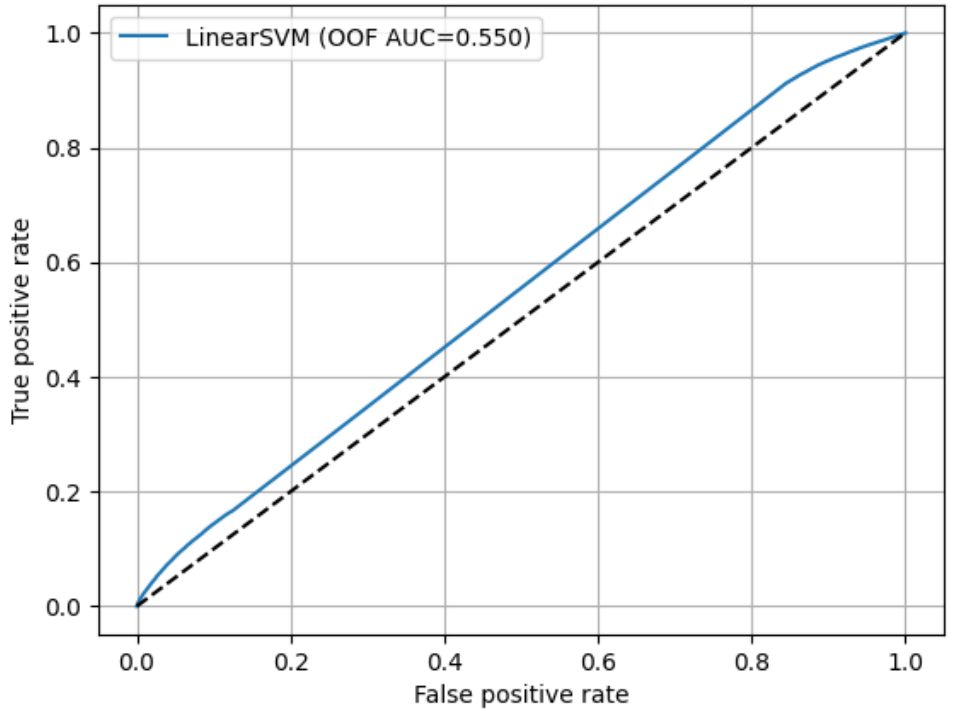
MODEL	AUC_CV_MEAN	AUC_Out of folds	BEST_PARAMS	WALLTIME_SEC
HistGBDT Histogram-based Gradient Boosting Classification	0.563974	0.56373	{'clf__min_samples_leaf': 20, 'clf__max_iter': 200, 'clf__max_depth': 3, 'clf__learning_rate': 0.1}	214.5186
RandomForest	0.56308	0.562716	{'clf__n_estimators': 300, 'clf__min_samples_leaf': 20, 'clf__max_features': 'sqrt', 'clf__max_depth': 8}	7295.539
MLP	0.561451	0.560534	{'clf__learning_rate_init': 0.003, 'clf__hidden_layer_sizes': (9, 4), 'clf__alpha': 0.01}	430.036
Logistic	0.55165	0.551344	{'clf__C': 0.1}	10.49122
Linear_SVM	0.549941	0.549682	{'clf__C': 10.0}	10.71546
LDA	0.547915	0.547607	{'clf__solver': 'svd'}	6.273678
NaiveBayes	0.54654	0.546426	{}	3.780756
KNN	0.523907	0.523906	{'clf__n_neighbors': 100, 'clf__weights': 'distance'}	7526.95

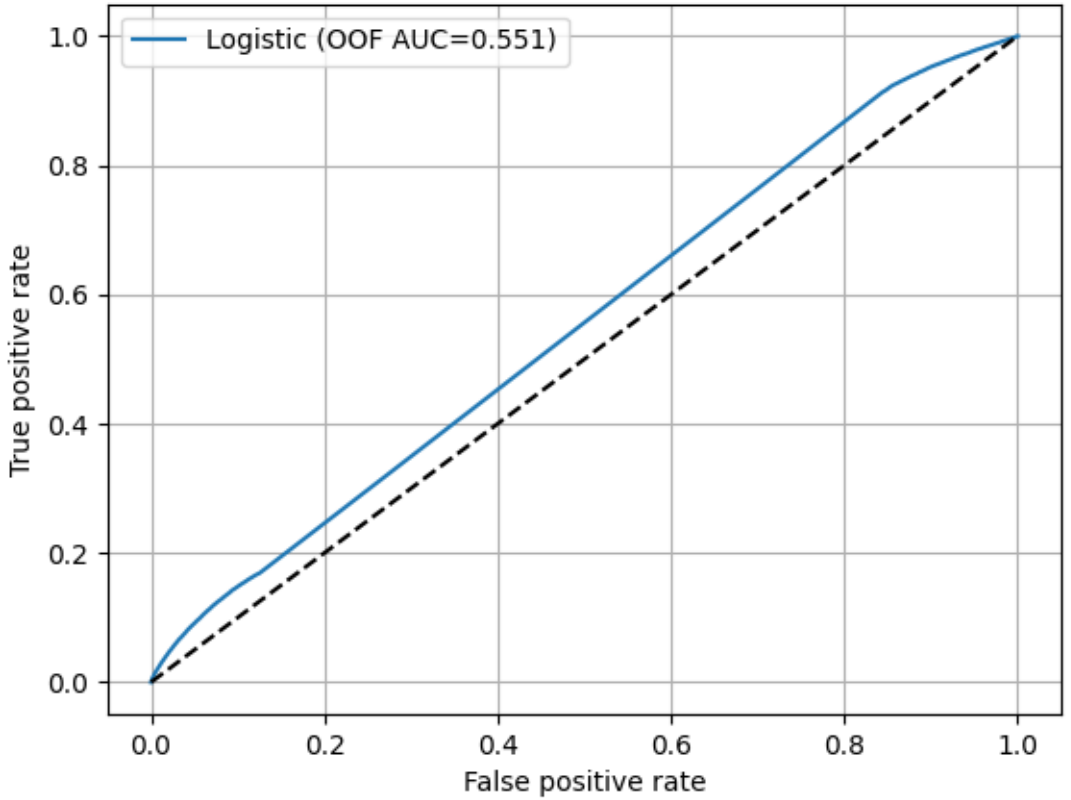
20.01.2026

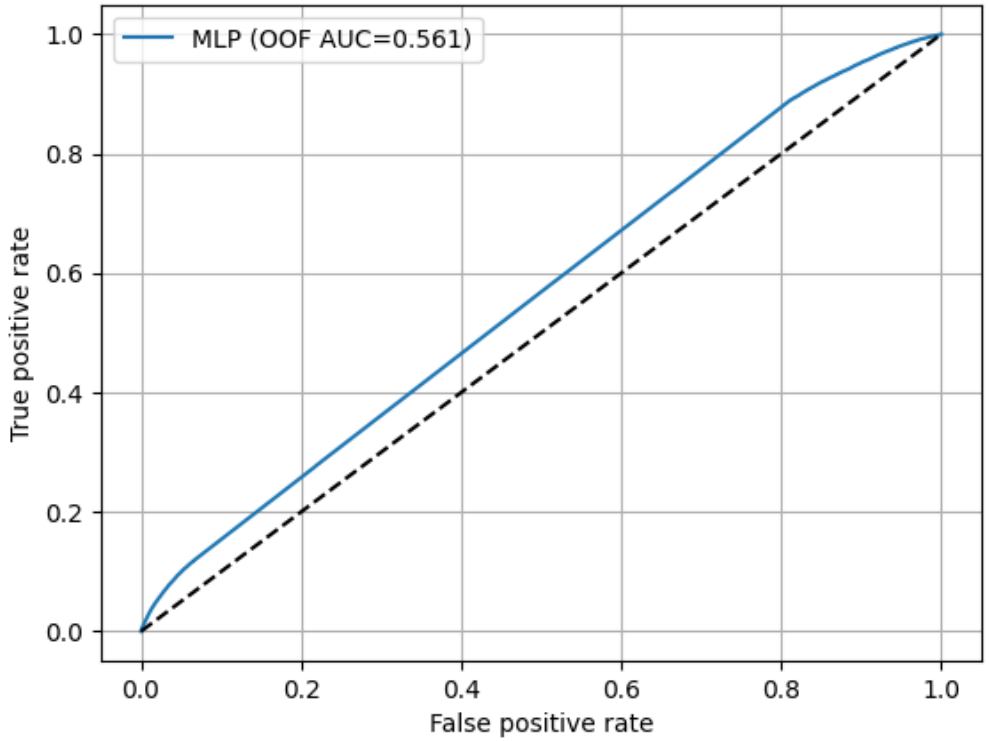

HistGBDT (OOF AUC=0.564)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

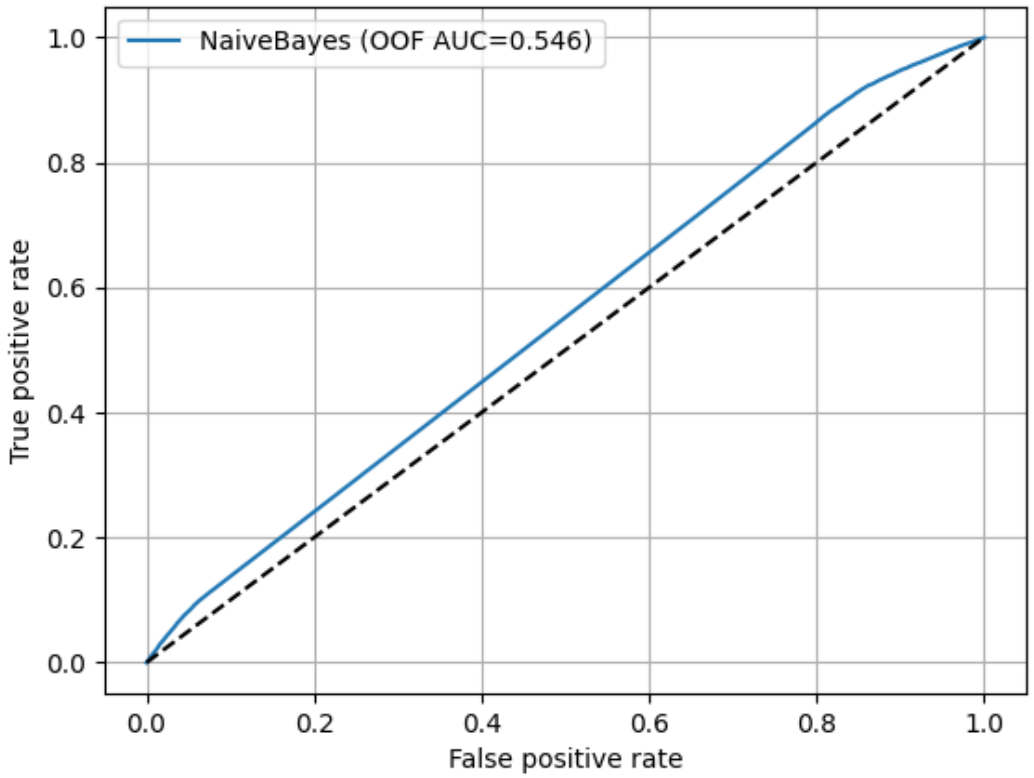

KNN (OOF AUC=0.524)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

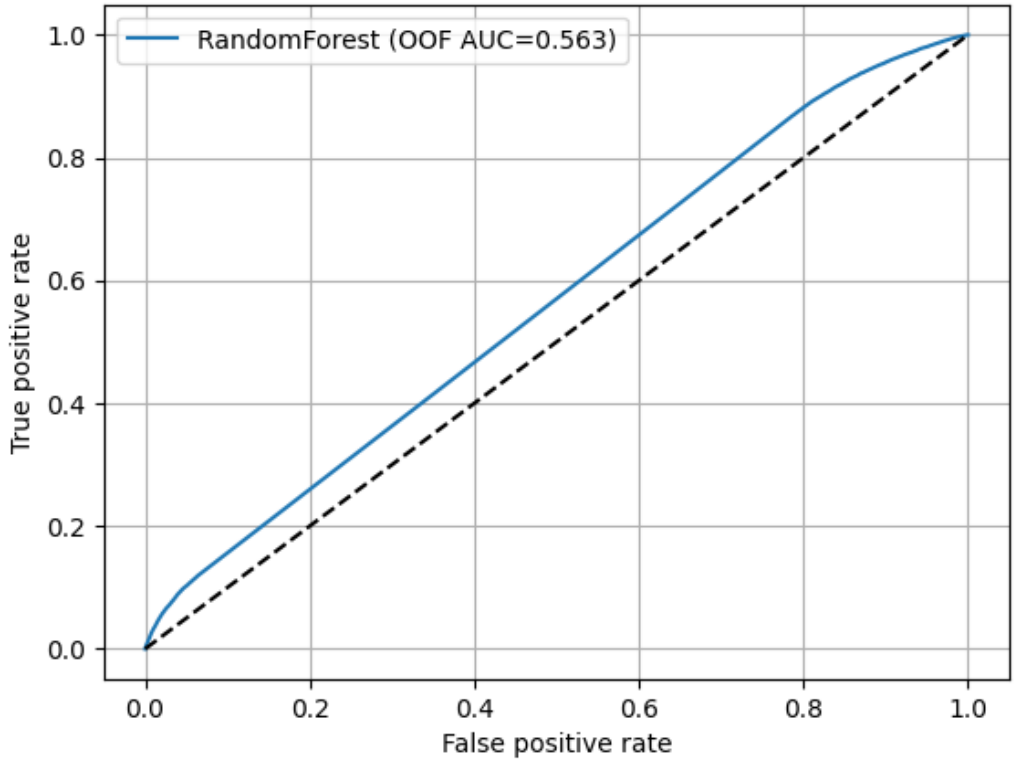

LDA (OOF AUC=0.548)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


LinearSVM (OOF AUC=0.550)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


Logistic (OOF AUC=0.551)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


MLP (OOF AUC=0.561)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


NaiveBayes (OOF AUC=0.546)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


RandomForest (OOF AUC=0.563)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

National Research
Tomsk
State
University

Лаборатория
анализа данных
физики высоких энергий
Томского
государственного
университета

Thank you for your attention!!!