XGBoost

Contents

1 Upload-Rate Prediction **2**
1.1 Reading the Data 2
1.2 Create the Prediction Tasks for Each Provider 3
1.3 Create Data Splitting Strategies for Testing and Validation 4
1.4 Create the Prediction Pipeline 5
1.5 Parameter Tuning 6
1.6 Create Learners with Tuned Hyperparameters 7
1.7 Validation Results 8
1.8 Feature Importance 12

2 Download-Rate Prediction **15**
2.1 Reading the Data 15
2.2 Create the Prediction Tasks for Each Provider 17
2.3 Parameter Tuning 17
2.4 Create Learners with Tuned Hyperparameters 18
2.5 Validation Results 18
2.6 Feature Importance 22

3 Link Lifetime Prediction **25**
3.1 Reading the Data 25
3.2 Create the Prediction Tasks for Each Provider 27
3.3 Parameter Tuning 27
3.4 Create Learners with Tuned Hyperparameters 28
3.5 Validation Results 28
3.6 Feature Importance 32

```r
library(tidyverse)
library(ggplot2)

library(mlr3)
library(mlr3learners)
library(mlr3pipelines)
library(mlr3tuning)
library(mlr3filters)
library(paradox)
```

```r
future::plan("multiprocess")
```

```
## Warning: Strategy 'multiprocess' is deprecated in future (>= 1.20.0). Instead,
## explicitly specify either 'multisession' or 'multicore'. In the current R
## session, 'multiprocess' equals 'multisession'.
```

1 Upload-Rate Prediction

1.1 Reading the Data

```r
data_dir = "../datasets/"
results_dir = "../prediction_results/"
```

```r
dataset_ul = read_csv(
  str_c(data_dir, "dataset_ul.csv"),
  col_types = cols(
    drive_id = col_integer(),
    scenario = col_factor(),
    provider = col_factor(),
    ci = col_factor(),
    enodeb = col_factor()
  )
) %>% select(
  drive_id,
  timestamp,
  scenario,
  provider,
  velocity_mps,
  rsrp_dbm,
  rsrq_db,
  rssnr_db,
  cqi,
  ta,
  enodeb,
  f_mhz,
  payload_mb,
  throughput_mbits
) %>% drop_na() %>% rowid_to_column(var="row_id_original")
```

```r
dataset_ul_o2 = filter(dataset_ul, provider=="o2")
glimpse(dataset_ul_o2)
```

```
## Rows: 2,039
## Columns: 15
## $ row_id_original  <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,...
## $ drive_id         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ timestamp        <dttm> 2018-12-10 09:08:57, 2018-12-10 09:09:08, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2...
## $ velocity_mps     <dbl> 11.80, 11.49, 7.93, 10.44, 10.92, 12.02, 10.28, 0....
## $ rsrp_dbm         <dbl> -99, -97, -96, -82, -101, -106, -112, -99, -98, -9...
## $ rsrq_db          <dbl> -9, -12, -12, -11, -14, -13, -18, -15, -15, -14, -...
## $ rssnr_db         <dbl> -1, -2, 5, 11, -3, -3, -6, -4, -6, -4, -6, -3, -2,...
## $ cqi              <dbl> 8, 9, 5, 15, 6, 6, 3, 4, 7, 4, 4, 5, 6, 5, 1, 4, 6...
## $ ta               <dbl> 9, 7, 7, 7, 7, 7, 7, 12, 13, 13, 13, 13, 11, 13, 1...
## $ enodeb           <fct> 54016, 52410, 52410, 52410, 52410, 52410, 52410, 5...
## $ f_mhz            <dbl> 1750, 1750, 1750, 1750, 1750, 1750, 1750, 880, 880...
## $ payload_mb       <dbl> 1.0, 6.0, 5.0, 7.0, 5.0, 8.0, 9.0, 7.0, 10.0, 2.0,...
## $ throughput_mbits <dbl> 4.66, 3.97, 6.52, 1.37, 0.80, 1.04, 2.34, 4.09, 2....
```

```
dataset_ul_tmobile = filter(dataset_ul, provider=="tmobile")
glimpse(dataset_ul_tmobile)
```

```
## Rows: 2,301
## Columns: 15
## $ row_id_original  <int> 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 20...
## $ drive_id         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ timestamp        <dttm> 2018-12-10 09:08:57, 2018-12-10 09:09:07, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> tmobile, tmobile, tmobile, tmobile, tmobile, tmobi...
## $ velocity_mps     <dbl> 11.83, 11.45, 8.15, 9.42, 10.61, 11.84, 9.75, 0.00...
## $ rsrp_dbm         <dbl> -85, -84, -74, -92, -90, -101, -93, -94, -94, -94,...
## $ rsrq_db          <dbl> -5, -6, -5, -6, -6, -10, -8, -11, -11, -10, -9, -1...
## $ rssnr_db         <dbl> 22, 11, 29, 13, 16, 13, 7, 0, 8, 2, 24, 10, 22, 15...
## $ cqi              <dbl> 10, 13, 15, 12, 9, 15, 10, 9, 9, 7, 10, 9, 12, 15,...
## $ ta               <dbl> 7, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 3, 3,...
## $ enodeb           <fct> 103068, 114809, 114809, 114809, 114809, 114809, 11...
## $ f_mhz            <dbl> 1720, 1720, 1720, 1720, 1720, 1720, 1720, 1720, 17...
## $ payload_mb       <dbl> 4.0, 2.0, 4.0, 9.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, ...
## $ throughput_mbits <dbl> 24.52, 14.86, 16.27, 12.68, 14.59, 13.13, 16.37, 1...
```

```
dataset_ul_vodafone = filter(dataset_ul, provider=="vodafone")
glimpse(dataset_ul_vodafone)
```

```
## Rows: 1,828
## Columns: 15
## $ row_id_original  <int> 4341, 4342, 4343, 4344, 4345, 4346, 4347, 4348, 43...
## $ drive_id         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ timestamp        <dttm> 2018-12-10 09:09:03, 2018-12-10 09:09:21, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> vodafone, vodafone, vodafone, vodafone, vodafone, ...
## $ velocity_mps     <dbl> 11.70, 8.22, 8.00, 10.30, 12.28, 0.00, 0.00, 0.00,...
## $ rsrp_dbm         <dbl> -121, -108, -111, -106, -110, -94, -95, -92, -98, ...
## $ rsrq_db          <dbl> -15, -9, -13, -8, -9, -7, -7, -8, -6, -10, -7, -8,...
## $ rssnr_db         <dbl> -8, 2, 6, 5, 9, 23, 23, 24, 14, 1, 14, 12, 14, 7, ...
## $ cqi              <dbl> 4, 2, 6, 11, 10, 15, 12, 15, 12, 6, 15, 10, 11, 7,...
## $ ta               <dbl> 63, 21, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16...
## $ enodeb           <fct> 51044, 52316, 50026, 50026, 50026, 50026, 50026, 5...
## $ f_mhz            <dbl> 1770, 1770, 1770, 1770, 1770, 1770, 1770, 1770, 17...
## $ payload_mb       <dbl> 6.0, 10.0, 0.1, 2.0, 6.0, 0.1, 0.1, 0.5, 7.0, 0.1,...
## $ throughput_mbits <dbl> 1.29, 3.18, 0.05, 2.93, 8.79, 5.16, 4.73, 10.13, 1...
```

1.2 Create the Prediction Tasks for Each Provider

```
make_task = function(dataset, task_id, target="throughput_mbits") {
  task = TaskRegr$new(
    id = task_id,
    backend = dataset %>% select(-drive_id, -timestamp, -provider, -scenario),
    target = target
  )

  task$col_roles$name = "row_id_original"
  task$col_roles$feature = setdiff(task$col_roles$feature, "row_id_original")
  return(task)
}

task_ul_o2 = make_task(dataset_ul_o2, "task_ul_o2")
task_ul_o2
```

```
## <TaskRegr:task_ul_o2> (2039 x 10)
## * Target: throughput_mbits
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, f_mhz, payload_mb, rsrp_dbm, rsrq_db, rssnr_db, ta,
##     velocity_mps
##   - fct (1): enodeb
```

```
task_ul_tmobile = make_task(dataset_ul_tmobile, "task_ul_tmobile")
task_ul_tmobile
```

```
## <TaskRegr:task_ul_tmobile> (2301 x 10)
## * Target: throughput_mbits
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, f_mhz, payload_mb, rsrp_dbm, rsrq_db, rssnr_db, ta,
##     velocity_mps
##   - fct (1): enodeb
```

```
task_ul_vodafone = make_task(dataset_ul_vodafone, "task_ul_vodafone")
task_ul_vodafone
```

```
## <TaskRegr:task_ul_vodafone> (1828 x 10)
## * Target: throughput_mbits
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, f_mhz, payload_mb, rsrp_dbm, rsrq_db, rssnr_db, ta,
##     velocity_mps
##   - fct (1): enodeb
```

1.3 Create Data Splitting Strategies for Testing and Validation

The outer resampling is used for the train/validation split.

```
get_row_ids_by_drive_ids = function(task, dataset, drive_ids) {
  result = (tibble(task$row_names) %>%
    inner_join(dataset, by=c("row_name"="row_id_original")) %>%
    filter(drive_id %in% drive_ids))$row_id
  return(result)
}

make_outer_resampling = function(task, dataset, drive_ids_train, drive_ids_test) {
  row_ids_train = get_row_ids_by_drive_ids(task, dataset, drive_ids_train)
  row_ids_test = get_row_ids_by_drive_ids(task, dataset, drive_ids_test)

  result = rsmp("custom")
  result$instantiate(task, train_sets=list(row_ids_train), test_sets=list(row_ids_test))

  return(result)
}
```

The inner resampling is used for the parameter tuning on the training set.

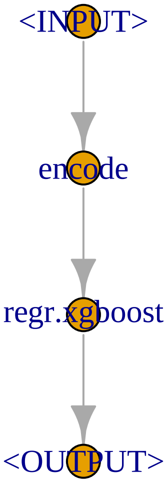
```r
make_inner_resampling = function(task, dataset, last_drive_id) {
  train_sets = list()
  test_sets = list()

  for (cur_last_drive_id_train in 2:(last_drive_id-1)) {
    drive_ids_train = 1:cur_last_drive_id_train
    drive_ids_test = cur_last_drive_id_train + 1

    row_ids_train = get_row_ids_by_drive_ids(task, dataset, drive_ids_train)
    row_ids_test = get_row_ids_by_drive_ids(task, dataset, drive_ids_test)

    train_sets[[length(train_sets)+1]] = row_ids_train
    test_sets[[length(test_sets)+1]] = row_ids_test
  }

  result = rsmp("custom")
  result$instantiate(task, train_sets=train_sets, test_sets=test_sets)

  return(result)
}
```

1.4 Create the Prediction Pipeline

```r
make_learner = function(nrounds=100, eta=NULL, gamma=NULL, lambda=NULL) {
  factor_encoding = po(
    "encode",
    method = "one-hot",
    affect_columns = selector_type("factor")
  )
  xgboost = lrn("regr.xgboost")

  if (!is.null(nrounds)) {
    xgboost$param_set$values = mlr3misc::insert_named(
      xgboost$param_set$values,
      list(nrounds=nrounds)
    )
  }
  if (!is.null(eta)) {
    xgboost$param_set$values = mlr3misc::insert_named(
      xgboost$param_set$values,
      list(eta=eta)
    )
  }
  if (!is.null(gamma)) {
    xgboost$param_set$values = mlr3misc::insert_named(
      xgboost$param_set$values,
      list(gamma=gamma)
    )
  }
  if (!is.null(lambda)) {
    xgboost$param_set$values = mlr3misc::insert_named(
      xgboost$param_set$values,
      list(lambda=lambda)
    )
  }

  pipe = factor_encoding %>>% PipeOpLearner$new(xgboost)
  learner = GraphLearner$new(pipe)
  return(learner)
}
```

Here we can see the prediction pipeline:

```
make_learner()$graph$plot()
```



1.5 Parameter Tuning

```
parameter_space = ParamSet$new(list(
  ParamInt$new("regr.xgboost.nrounds", lower=100, upper=1000),
  ParamDbl$new("regr.xgboost.eta", lower=0.01, upper=1),
  ParamDbl$new("regr.xgboost.gamma", lower=0, upper=10),
  ParamDbl$new("regr.xgboost.lambda", lower=0, upper=10)
))
```

```
get_tuning_result = function(task, dataset, grid_resolution, n_evals) {
  tuning_instance = TuningInstanceSingleCrit$new(
    task = task,
    learner = make_learner(),
    resampling = make_inner_resampling(task, dataset, last_drive_id=7),
    measure = msr("regr.mae"),
    terminator = trm("evals", n_evals=n_evals),
    search_space = parameter_space$clone(deep = TRUE),
    store_benchmark_result = TRUE,
    check_values = TRUE
  )

  tuner = tnr("grid_search", resolution = grid_resolution)
  tuner$optimize(tuning_instance)

  return(tuning_instance)
}
```

```
tuning_result_ul_o2 = get_tuning_result(task_ul_o2, dataset_ul, grid_resolution = 20, n_evals = 10)

tuning_result_ul_tmobile = get_tuning_result(task_ul_tmobile, dataset_ul, grid_resolution = 20, n_evals

tuning_result_ul_vodafone = get_tuning_result(task_ul_vodafone, dataset_ul, grid_resolution = 20, n_eva
```

```
tuning_result_ul = bind_rows(
  tibble(tuning_result_ul_o2$result) %>% mutate(provider="o2"),
  tibble(tuning_result_ul_tmobile$result) %>% mutate(provider="tmobile"),
  tibble(tuning_result_ul_vodafone$result) %>% mutate(provider="vodafone"),
) %>% select("provider", "regr.xgboost.nrounds", "regr.xgboost.eta", "regr.xgboost.gamma", "regr.xgboos

knitr::kable(tuning_result_ul)
```

provider	regr.xgboost.nrounds	regr.xgboost.eta	regr.xgboost.gamma	regr.xgboost.lambda
o2	1000	0.1663158	7.894737	7.368421
tmobile	526	0.1142105	7.368421	2.631579
vodafone	289	0.0621053	2.631579	10.000000

1.6 Create Learners with Tuned Hyperparameters

```
learner_ul_o2 = make_learner(
  nrounds = tuning_result_ul_o2$result$regr.xgboost.nrounds,
  eta = tuning_result_ul_o2$result$regr.xgboost.eta,
  gamma = tuning_result_ul_o2$result$regr.xgboost.gamma,
  lambda = tuning_result_ul_o2$result$regr.xgboost.lambda
)

learner_ul_tmobile = make_learner(
  nrounds = tuning_result_ul_tmobile$result$regr.xgboost.nrounds,
  eta = tuning_result_ul_tmobile$result$regr.xgboost.eta,
  gamma = tuning_result_ul_tmobile$result$regr.xgboost.gamma,
  lambda = tuning_result_ul_tmobile$result$regr.xgboost.lambda
)

learner_ul_vodafone = make_learner(
  nrounds = tuning_result_ul_vodafone$result$regr.xgboost.nrounds,
  eta = tuning_result_ul_vodafone$result$regr.xgboost.eta,
  gamma = tuning_result_ul_vodafone$result$regr.xgboost.gamma,
  lambda = tuning_result_ul_vodafone$result$regr.xgboost.lambda
)
```

1.7 Validation Results

```
resampling_result_ul_o2 = resample(
  task = task_ul_o2,
  learner = learner_ul_o2,
  resampling = make_outer_resampling(task_ul_o2, dataset_ul, drive_ids_train=1:7, drive_ids_test=8:10),
  store_models = TRUE
)

resampling_result_ul_tmobile = resample(
  task = task_ul_tmobile,
  learner = learner_ul_tmobile,
  resampling = make_outer_resampling(task_ul_tmobile, dataset_ul, drive_ids_train=1:7, drive_ids_test=8
  store_models = TRUE
)

resampling_result_ul_vodafone = resample(
  task = task_ul_vodafone,
  learner = learner_ul_vodafone,
  resampling = make_outer_resampling(task_ul_vodafone, dataset_ul, drive_ids_train=1:7, drive_ids_test=8
  store_models = TRUE
)
```

```
predictions_ul_o2 = as.data.table(resampling_result_ul_o2$prediction())
glimpse(tibble(predictions_ul_o2))
```

```
## Rows: 615
## Columns: 3
## $ row_id   <int> 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148...
## $ truth    <dbl> 4.47, 2.59, 2.26, 1.09, 0.77, 0.19, 0.26, 0.65, 1.45, 1.12...
## $ response <dbl> 3.4680877, 2.3749433, 3.8357697, 2.9707620, 3.4179711, 1.1...
```

```
predictions_ul_tmobile = as.data.table(resampling_result_ul_tmobile$prediction())
predictions_ul_vodafone = as.data.table(resampling_result_ul_vodafone$prediction())
```

```
validation_results_ul = bind_rows(
  tibble(predictions_ul_o2) %>%
    inner_join(tibble(task_ul_o2$row_names), by="row_id") %>%
    inner_join(dataset_ul, by=c("row_name"="row_id_original")),
  tibble(predictions_ul_tmobile) %>%
    inner_join(tibble(task_ul_tmobile$row_names), by="row_id") %>%
    inner_join(dataset_ul, by=c("row_name"="row_id_original")),
  tibble(predictions_ul_vodafone) %>%
    inner_join(tibble(task_ul_vodafone$row_names), by="row_id") %>%
    inner_join(dataset_ul, by=c("row_name"="row_id_original"))
)
glimpse(validation_results_ul)
```

```
## Rows: 1,840
## Columns: 18
## $ row_id           <int> 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, ...
## $ truth            <dbl> 4.47, 2.59, 2.26, 1.09, 0.77, 0.19, 0.26, 0.65, 1....
## $ response         <dbl> 3.4680877, 2.3749433, 3.8357697, 2.9707620, 3.4179...
## $ row_name         <int> 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, ...
## $ drive_id         <int> 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9,...
## $ timestamp        <dttm> 2018-12-11 09:04:11, 2018-12-11 09:04:22, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2...
## $ velocity_mps     <dbl> 0.00, 6.11, 9.39, 8.45, 11.68, 0.00, 0.00, 0.00, 4...
## $ rsrp_dbm         <dbl> -89, -92, -94, -98, -102, -100, -101, -101, -100, ...
## $ rsrq_db          <dbl> -9, -12, -14, -15, -16, -17, -16, -16, -17, -14, -...
## $ rssnr_db         <dbl> 13, 3, -1, -3, -5, -7, -6, -5, -8, 1, -7, -1, -2, ...
## $ cqi              <dbl> 11, 5, 5, 4, 2, 3, 4, 4, 4, 6, 3, 5, 5, 2, 4, 6, 3...
## $ ta               <dbl> 7, 7, 7, 7, 7, 12, 12, 12, 12, 12, 12, 12, 12, 12,...
## $ enodeb           <fct> 52410, 52410, 52410, 52410, 52410, 52900, 52900, 5...
## $ f_mhz            <dbl> 880, 880, 880, 880, 880, 880, 880, 880, 880, 880, ...
## $ payload_mb       <dbl> 0.1, 0.5, 3.0, 9.0, 7.0, 3.0, 2.0, 2.0, 6.0, 3.0, ...
## $ throughput_mbits <dbl> 4.47, 2.59, 2.26, 1.09, 0.77, 0.19, 0.26, 0.65, 1....
```

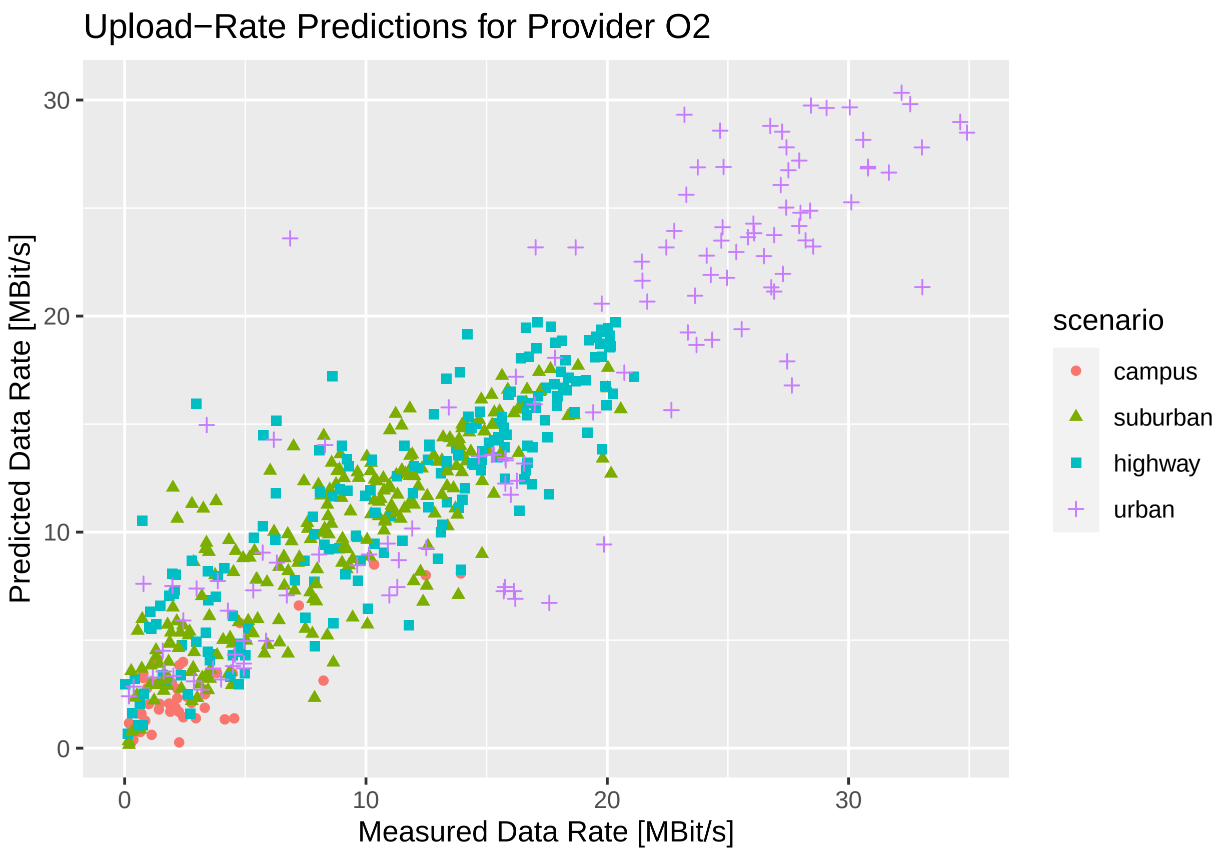
```r
all(validation_results_ul$truth == validation_results_ul$throughput_mbits)
```

```
## [1] TRUE
```

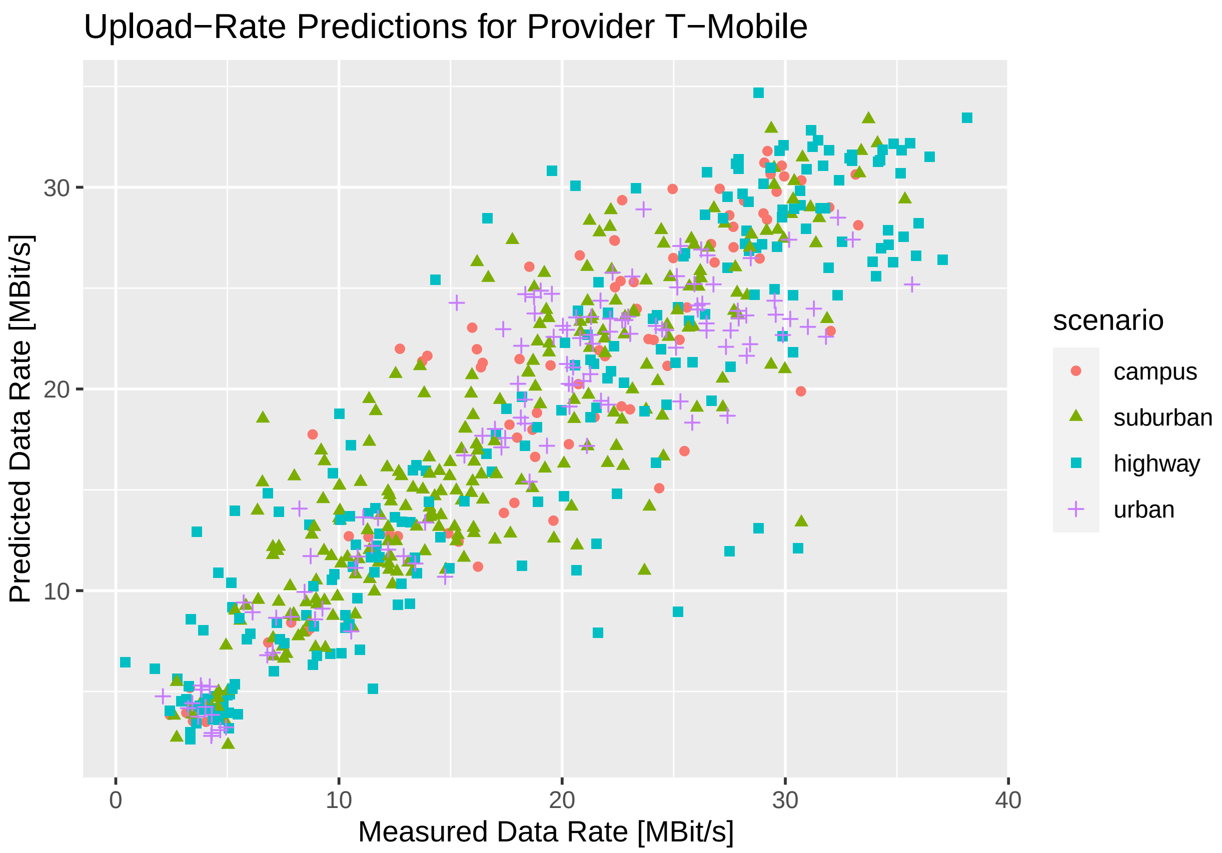
```r
validation_results_ul = validation_results_ul %>%
  rename(prediction_xgboost=response) %>%
  select(-truth, -row_id, -row_name)

# write_csv(validation_results_ul, str_c(results_dir, "predictions_xgboost_ul.csv"))
```

1.7.1 Scatter Plots

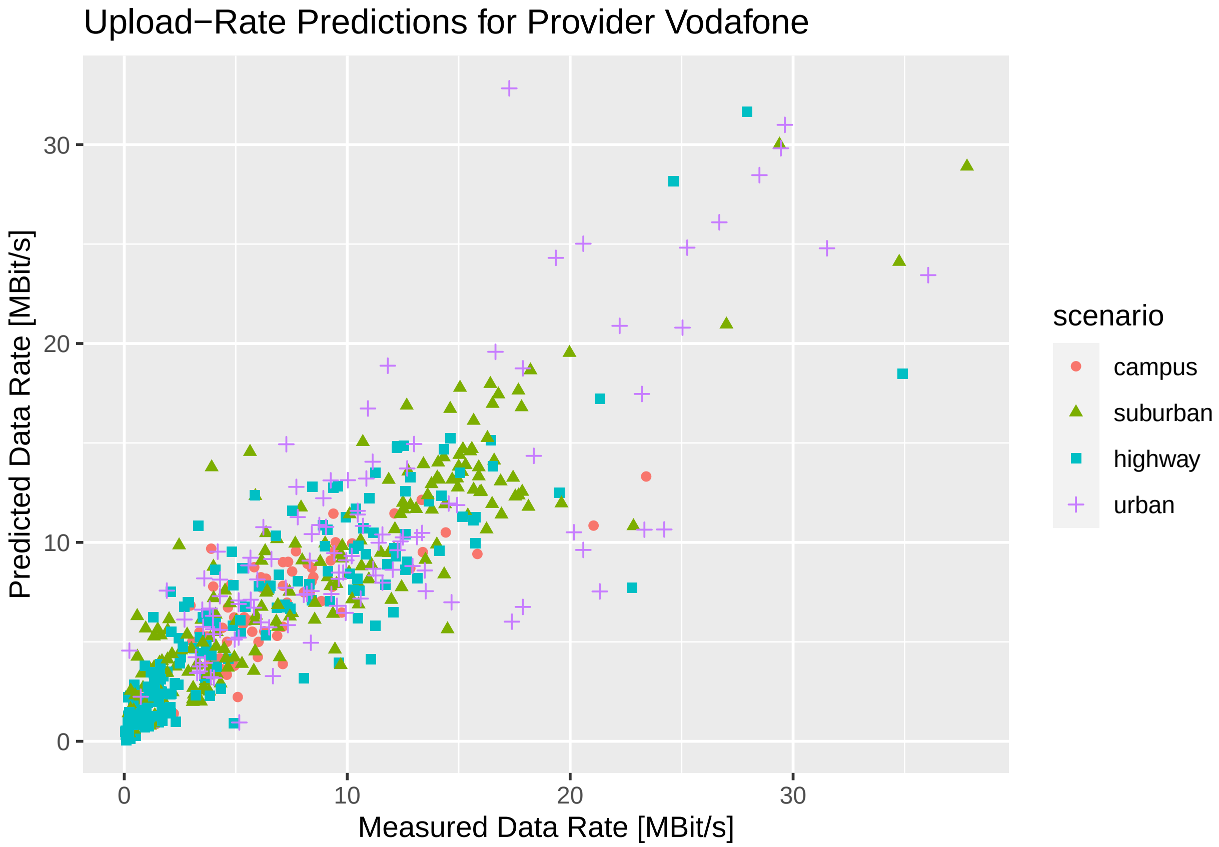
```r
ggplot(filter(validation_results_ul, provider=="o2"), aes(x=throughput_mbits, y=prediction_xgboost)) +
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Data Rate [MBit/s]") +
  ylab("Predicted Data Rate [MBit/s]") +
  ggtitle("Upload-Rate Predictions for Provider O2")
```



```
ggplot(filter(validation_results_ul, provider=="tmobile"), aes(x=throughput_mbits, y=prediction_xgboost
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Data Rate [MBit/s]") +
  ylab("Predicted Data Rate [MBit/s]") +
  ggtitle("Upload-Rate Predictions for Provider T-Mobile")
```



```
ggplot(filter(validation_results_ul, provider=="vodafone"), aes(x=throughput_mbits, y=prediction_xgboos
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Data Rate [MBit/s]") +
  ylab("Predicted Data Rate [MBit/s]") +
  ggtitle("Upload-Rate Predictions for Provider Vodafone")
```



1.8 Feature Importance

1.8.1 XGBoost Gain

```
importance_ul_o2 = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_ul_o2$learners[[1]]$model$regr.xgboost$model$feature_names,
  model = resampling_result_ul_o2$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "o2")
```

```
## [21:01:52] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecated
```

```
importance_ul_tmobile = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_ul_tmobile$learners[[1]]$model$regr.xgboost$model$feature_names,
  model = resampling_result_ul_tmobile$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "tmobile")
```

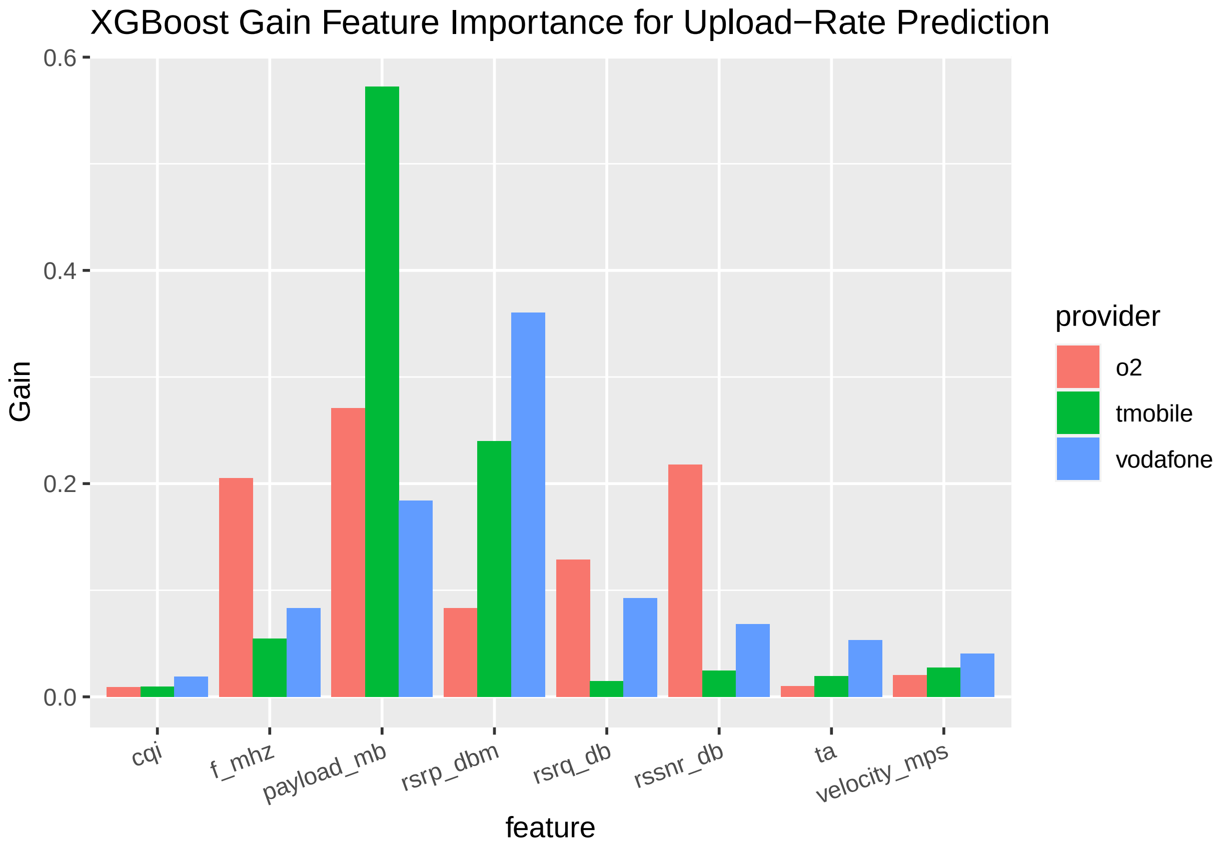
```
## [21:01:52] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecated
```

```
importance_ul_vodafone = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_ul_vodafone$learners[[1]]$model$regr.xgboost$model$feature_names,
  model = resampling_result_ul_vodafone$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "vodafone")
```

```
## [21:01:53] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecated
```

```
importance_ul = bind_rows(
  importance_ul_o2,
  importance_ul_tmobile,
  importance_ul_vodafone
) %>% filter(!str_starts(Feature, "enodeb"))
```

```
ggplot(importance_ul) +
  geom_bar(position = "dodge", aes(x = Feature, y = Gain, fill = provider), stat="identity") +
  xlab("feature") +
  ylab("Gain") +
  scale_x_discrete(guide = guide_axis(angle = 20)) +
  ggtitle("XGBoost Gain Feature Importance for Upload-Rate Prediction")
```

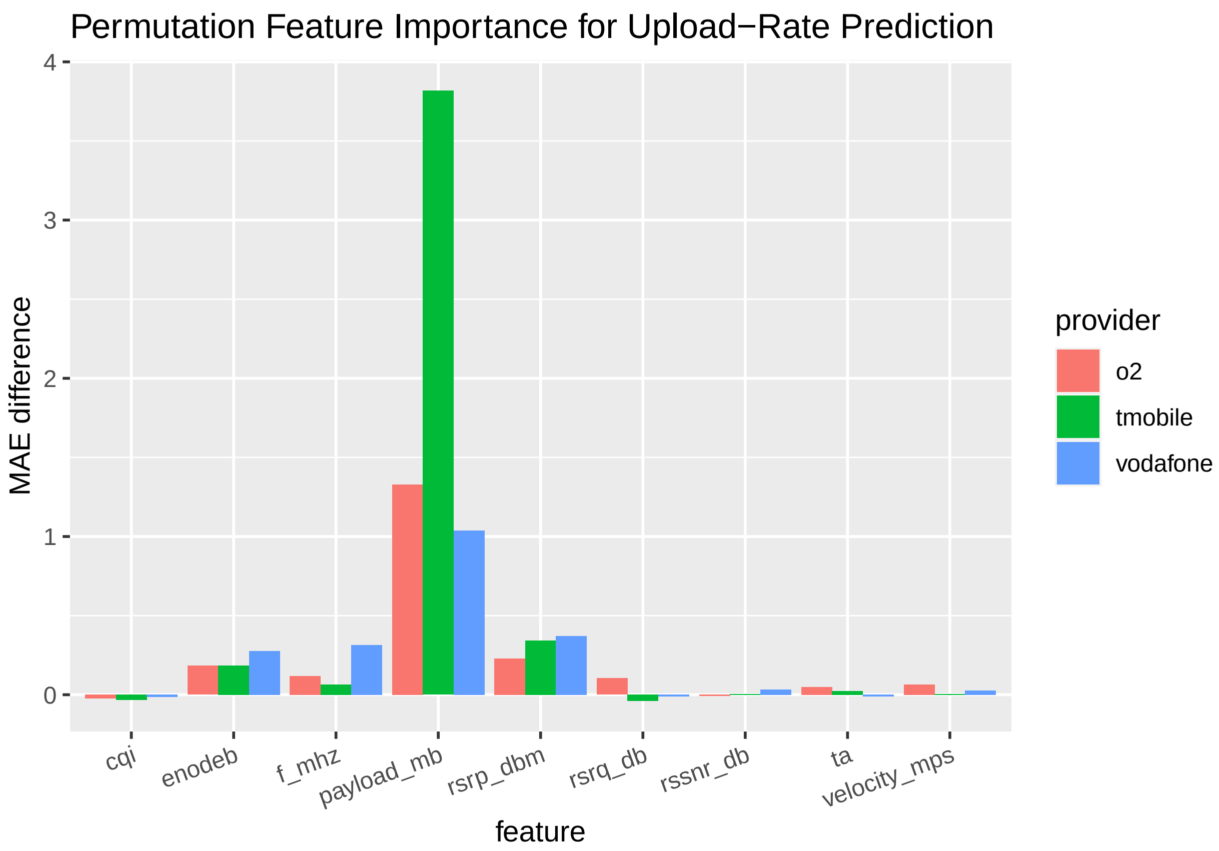


1.8.2 Permutation

```
uninstantiate_resampling = function(resampling) {
  new_resampling = new.env()
  class(new_resampling) = class(resampling)
  for (val in ls(resampling, all.names = TRUE)) {
    if (val != "is_instantiated") {
      assign(val, get(val, envir=resampling), envir = new_resampling)
    }
  }
  new_resampling$is_instantiated = FALSE

  return(new_resampling)
}
```

```r
num_permutation_sims_ul = 1

filter_permutation_o2_ul = flt("permutation",
  learner = learner_ul_o2$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_ul_o2, dataset_ul, drive_ids_train=1:7, drive_ids_test=8:10)
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc=num_permutation_sims_ul
)
filter_permutation_o2_ul$calculate(task_ul_o2)
permutation_ul_o2 = tibble(as.data.table(filter_permutation_o2_ul)) %>% mutate(provider="o2")

filter_permutation_tmobile_ul = flt("permutation",
  learner = learner_ul_tmobile$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_ul_tmobile, dataset_ul, drive_ids_train=1:7, drive_ids_test=8:10)
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc=num_permutation_sims_ul
)
filter_permutation_tmobile_ul$calculate(task_ul_tmobile)
permutation_ul_tmobile = tibble(as.data.table(filter_permutation_tmobile_ul)) %>% mutate(provider="tmobi

filter_permutation_vodafone_ul = flt("permutation",
  learner = learner_ul_vodafone$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_ul_vodafone, dataset_ul, drive_ids_train=1:7, drive_ids_test=8:10)
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc=num_permutation_sims_ul
)
filter_permutation_vodafone_ul$calculate(task_ul_vodafone)
permutation_ul_vodafone = tibble(as.data.table(filter_permutation_vodafone_ul)) %>% mutate(provider="vo

permutation_ul = bind_rows(
  permutation_ul_o2,
  permutation_ul_tmobile,
  permutation_ul_vodafone
)
```

```r
ggplot(permutation_ul) +
  geom_bar(position = "dodge", aes(x = feature, y = score, fill = provider), stat="identity") +
  xlab("feature") +
  ylab("MAE difference") +
  scale_x_discrete(guide = guide_axis(angle = 20)) +
  ggtitle("Permutation Feature Importance for Upload-Rate Prediction")
```



2 Download-Rate Prediction

2.1 Reading the Data

```
dataset_dl = read_csv(
  str_c(data_dir, "dataset_dl.csv"),
  col_types = cols(
    drive_id = col_integer(),
    scenario = col_factor(),
    provider = col_factor(),
    ci = col_factor(),
    enodeb = col_factor()
  )
) %>% select(
  drive_id,
  timestamp,
  scenario,
  provider,
  velocity_mps,
  rsrp_dbm,
  rsrq_db,
  rssnr_db,
  cqi,
  ta,
  enodeb,
  f_mhz,
  payload_mb,
  throughput_mbits
) %>% drop_na() %>% rowid_to_column(var="row_id_original")
```

```
dataset_dl_o2 = filter(dataset_dl, provider=="o2")
glimpse(dataset_dl_o2)
```

```
## Rows: 2,033
## Columns: 15
## $ row_id_original  <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,...
## $ drive_id         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ timestamp        <dttm> 2018-12-10 09:08:56, 2018-12-10 09:09:15, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2...
## $ velocity_mps     <dbl> 11.80, 8.02, 7.93, 10.08, 12.44, 12.03, 11.52, 0.0...
## $ rsrp_dbm         <dbl> -99, -96, -96, -100, -101, -105, -112, -99, -98, -...
## $ rsrq_db          <dbl> -9, -12, -12, -13, -14, -15, -18, -15, -13, -14, -...
## $ rssnr_db         <dbl> -1, 5, 5, -1, -3, -4, -6, -4, -6, -4, -6, -3, -2, ...
## $ cqi              <dbl> 8, 5, 5, 7, 5, 5, 3, 4, 6, 4, 5, 5, 6, 5, 1, 4, 6,...
## $ ta               <dbl> 9, 7, 7, 7, 7, 7, 7, 12, 13, 13, 13, 13, 11, 13, 1...
## $ enodeb           <fct> 54016, 52410, 52410, 52410, 52410, 52410, 52410, 5...
## $ f_mhz            <dbl> 1845, 1845, 1845, 1845, 1845, 1845, 1845, 850, 850...
## $ payload_mb       <dbl> 6.0, 10.0, 7.0, 7.0, 9.0, 3.0, 3.0, 0.5, 5.0, 2.0,...
## $ throughput_mbits <dbl> 2.38, 0.90, 1.09, 0.45, 0.51, 0.42, 0.71, 0.63, 0....
```

```
dataset_dl_tmobile = filter(dataset_dl, provider=="tmobile")
glimpse(dataset_dl_tmobile)
```

```
## Rows: 2,300
## Columns: 15
## $ row_id_original  <int> 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 20...
## $ drive_id         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ timestamp        <dttm> 2018-12-10 09:08:57, 2018-12-10 09:09:08, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> tmobile, tmobile, tmobile, tmobile, tmobile, tmobi...
## $ velocity_mps     <dbl> 11.83, 11.45, 8.15, 9.42, 10.61, 11.84, 9.75, 0.00...
## $ rsrp_dbm         <dbl> -85, -84, -74, -92, -90, -101, -93, -94, -94, -94,...
## $ rsrq_db          <dbl> -5, -6, -5, -6, -6, -10, -8, -11, -11, -10, -9, -1...
## $ rssnr_db         <dbl> 22, 11, 29, 13, 16, 13, 7, 0, 8, 2, 24, 10, 22, 15...
## $ cqi              <dbl> 10, 13, 15, 12, 9, 15, 10, 9, 9, 7, 10, 9, 12, 15,...
## $ ta               <dbl> 7, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 3, 3,...
## $ enodeb           <fct> 103068, 114809, 114809, 114809, 114809, 114809, 11...
## $ f_mhz            <dbl> 1815, 1815, 1815, 1815, 1815, 1815, 1815, 1815, 18...
## $ payload_mb       <dbl> 0.1, 2.0, 2.0, 3.0, 5.0, 8.0, 3.0, 0.1, 5.0, 9.0, ...
## $ throughput_mbits <dbl> 6.84, 9.71, 7.31, 3.95, 8.55, 11.30, 6.78, 4.30, 1...
```

```
dataset_dl_vodafone = filter(dataset_dl, provider=="vodafone")
glimpse(dataset_dl_vodafone)
```

```
## Rows: 2,170
## Columns: 15
## $ row_id_original  <int> 4334, 4335, 4336, 4337, 4338, 4339, 4340, 4341, 43...
## $ drive_id         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ timestamp        <dttm> 2018-12-10 09:09:03, 2018-12-10 09:09:21, 2018-12...
## $ scenario         <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider         <fct> vodafone, vodafone, vodafone, vodafone, vodafone, ...
## $ velocity_mps     <dbl> 11.70, 8.22, 8.00, 10.60, 10.30, 12.28, 11.45, 0.0...
## $ rsrp_dbm         <dbl> -121, -108, -111, -113, -106, -110, -93, -94, -95,...
## $ rsrq_db          <dbl> -15, -9, -13, -11, -8, -9, -5, -7, -7, -8, -6, -6,...
## $ rssnr_db         <dbl> -8, 2, 6, 1, 5, 9, 21, 23, 23, 24, 14, 23, 13, 1, ...
## $ cqi              <dbl> 4, 2, 6, 6, 11, 10, 14, 15, 12, 15, 12, 14, 15, 6,...
## $ ta               <dbl> 63, 21, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16...
## $ enodeb           <fct> 51044, 52316, 50026, 50026, 50026, 50026, 50026, 5...
## $ f_mhz            <dbl> 1865, 1865, 1865, 1865, 1865, 1865, 1865, 1865, 18...
## $ payload_mb       <dbl> 0.1, 5.0, 1.0, 3.0, 8.0, 4.0, 0.5, 5.0, 6.0, 0.5, ...
## $ throughput_mbits <dbl> 3.54, 18.57, 5.22, 3.97, 11.68, 35.91, 25.32, 62.7...
```

2.2 Create the Prediction Tasks for Each Provider

```
task_dl_o2 = make_task(dataset_dl_o2, "task_dl_o2")
task_dl_o2
```

```
## <TaskRegr:task_dl_o2> (2033 x 10)
## * Target: throughput_mbits
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, f_mhz, payload_mb, rsrp_dbm, rsrq_db, rssnr_db, ta,
##     velocity_mps
##   - fct (1): enodeb
```

```
task_dl_tmobile = make_task(dataset_dl_tmobile, "task_dl_tmobile")
task_dl_tmobile
```

```
## <TaskRegr:task_dl_tmobile> (2300 x 10)
## * Target: throughput_mbits
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, f_mhz, payload_mb, rsrp_dbm, rsrq_db, rssnr_db, ta,
##     velocity_mps
##   - fct (1): enodeb
```

```
task_dl_vodafone = make_task(dataset_dl_vodafone, "task_dl_vodafone")
task_dl_vodafone
```

```
## <TaskRegr:task_dl_vodafone> (2170 x 10)
## * Target: throughput_mbits
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, f_mhz, payload_mb, rsrp_dbm, rsrq_db, rssnr_db, ta,
##     velocity_mps
##   - fct (1): enodeb
```

2.3 Parameter Tuning

```
tuning_result_dl_o2 = get_tuning_result(task_dl_o2, dataset_dl, grid_resolution = 20, n_evals = 10)

tuning_result_dl_tmobile = get_tuning_result(task_dl_tmobile, dataset_dl, grid_resolution = 20, n_evals

tuning_result_dl_vodafone = get_tuning_result(task_dl_vodafone, dataset_dl, grid_resolution = 20, n_eva
```

```r
tuning_result_dl = bind_rows(
  tibble(tuning_result_dl_o2$result) %>% mutate(provider="o2"),
  tibble(tuning_result_dl_tmobile$result) %>% mutate(provider="tmobile"),
  tibble(tuning_result_dl_vodafone$result) %>% mutate(provider="vodafone"),
) %>% select("provider", "regr.xgboost.nrounds", "regr.xgboost.eta", "regr.xgboost.gamma", "regr.xgboos

knitr::kable(tuning_result_dl)
```

provider	regr.xgboost.nrounds	regr.xgboost.eta	regr.xgboost.gamma	regr.xgboost.lambda
o2	1000	0.3226316	6.315790	0.0000000
tmobile	621	0.1142105	8.947368	0.5263158
vodafone	479	0.3226316	4.736842	7.8947368

2.4 Create Learners with Tuned Hyperparameters

```r
learner_dl_o2 = make_learner(
  nrounds = tuning_result_dl_o2$result$regr.xgboost.nrounds,
  eta = tuning_result_dl_o2$result$regr.xgboost.eta,
  gamma = tuning_result_dl_o2$result$regr.xgboost.gamma,
  lambda = tuning_result_dl_o2$result$regr.xgboost.lambda
)

learner_dl_tmobile = make_learner(
  nrounds = tuning_result_dl_tmobile$result$regr.xgboost.nrounds,
  eta = tuning_result_dl_tmobile$result$regr.xgboost.eta,
  gamma = tuning_result_dl_tmobile$result$regr.xgboost.gamma,
  lambda = tuning_result_dl_tmobile$result$regr.xgboost.lambda
)

learner_dl_vodafone = make_learner(
  nrounds = tuning_result_dl_vodafone$result$regr.xgboost.nrounds,
  eta = tuning_result_dl_vodafone$result$regr.xgboost.eta,
  gamma = tuning_result_dl_vodafone$result$regr.xgboost.gamma,
  lambda = tuning_result_dl_vodafone$result$regr.xgboost.lambda
)
```

2.5 Validation Results

```r
resampling_result_dl_o2 = resample(
  task = task_dl_o2,
  learner = learner_dl_o2,
  resampling = make_outer_resampling(task_dl_o2, dataset_dl, drive_ids_train=1:7, drive_ids_test=8:10),
  store_models = TRUE
)

resampling_result_dl_tmobile = resample(
  task = task_dl_tmobile,
  learner = learner_dl_tmobile,
  resampling = make_outer_resampling(task_dl_tmobile, dataset_dl, drive_ids_train=1:7, drive_ids_test=8
  store_models = TRUE
)
```

```r
resampling_result_dl_vodafone = resample(
  task = task_dl_vodafone,
  learner = learner_dl_vodafone,
  resampling = make_outer_resampling(task_dl_vodafone, dataset_dl, drive_ids_train=1:7, drive_ids_test=8
  store_models = TRUE
)
```

```r
predictions_dl_o2 = as.data.table(resampling_result_dl_o2$prediction())
glimpse(tibble(predictions_dl_o2))
```

```
## Rows: 609
## Columns: 3
## $ row_id   <int> 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157...
## $ truth    <dbl> 3.72, 2.56, 0.69, 0.36, 0.55, 0.36, 0.37, 0.28, 0.12, 1.36...
## $ response <dbl> 7.83577538, 0.96934938, 1.91151321, 0.62662381, 1.08720112...
```

```r
predictions_dl_tmobile = as.data.table(resampling_result_dl_tmobile$prediction())
predictions_dl_vodafone = as.data.table(resampling_result_dl_vodafone$prediction())
```

```r
validation_results_dl = bind_rows(
  tibble(predictions_dl_o2) %>%
    inner_join(tibble(task_dl_o2$row_names), by="row_id") %>%
    inner_join(dataset_dl, by=c("row_name"="row_id_original")),
  tibble(predictions_dl_tmobile) %>%
    inner_join(tibble(task_dl_tmobile$row_names), by="row_id") %>%
    inner_join(dataset_dl, by=c("row_name"="row_id_original")),
  tibble(predictions_dl_vodafone) %>%
    inner_join(tibble(task_dl_vodafone$row_names), by="row_id") %>%
    inner_join(dataset_dl, by=c("row_name"="row_id_original"))
)
glimpse(validation_results_dl)
```

```
## Rows: 1,923
## Columns: 18
## $ row_id          <int> 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, ...
## $ truth           <dbl> 3.72, 2.56, 0.69, 0.36, 0.55, 0.36, 0.37, 0.28, 0....
## $ response        <dbl> 7.83577538, 0.96934938, 1.91151321, 0.62662381, 1....
## $ row_name        <int> 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, ...
## $ drive_id        <int> 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 9, 9,...
## $ timestamp       <dttm> 2018-12-11 09:04:12, 2018-12-11 09:04:23, 2018-12...
## $ scenario        <fct> campus, campus, campus, campus, campus, campus, ca...
## $ provider        <fct> o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2...
## $ velocity_mps    <dbl> 0.00, 7.20, 8.74, 9.01, 9.00, 11.68, 0.00, 0.00, 1...
## $ rsrp_dbm        <dbl> -89, -92, -88, -99, -105, -102, -100, -101, -100, ...
## $ rsrq_db         <dbl> -9, -12, -12, -15, -16, -16, -17, -16, -14, -16, -...
## $ rssnr_db        <dbl> 13, 3, 7, -4, -3, -5, -7, -5, 1, -4, -7, -6, 0, -2...
## $ cqi             <dbl> 11, 5, 6, 4, 3, 2, 3, 4, 6, 5, 3, 4, 5, 6, 5, 6, 4...
## $ ta              <dbl> 7, 7, 7, 7, 7, 7, 12, 12, 12, 7, 12, 12, 12, 12, 1...
## $ enodeb          <fct> 52410, 52410, 52410, 52410, 52410, 52410, 52900, 5...
## $ f_mhz           <dbl> 850, 850, 850, 850, 850, 850, 850, 850, 850, 850, ...
## $ payload_mb      <dbl> 6.0, 2.0, 7.0, 2.0, 8.0, 2.0, 8.0, 6.0, 0.5, 1.0, ...
## $ throughput_mbits <dbl> 3.72, 2.56, 0.69, 0.36, 0.55, 0.36, 0.37, 0.28, 0....
```

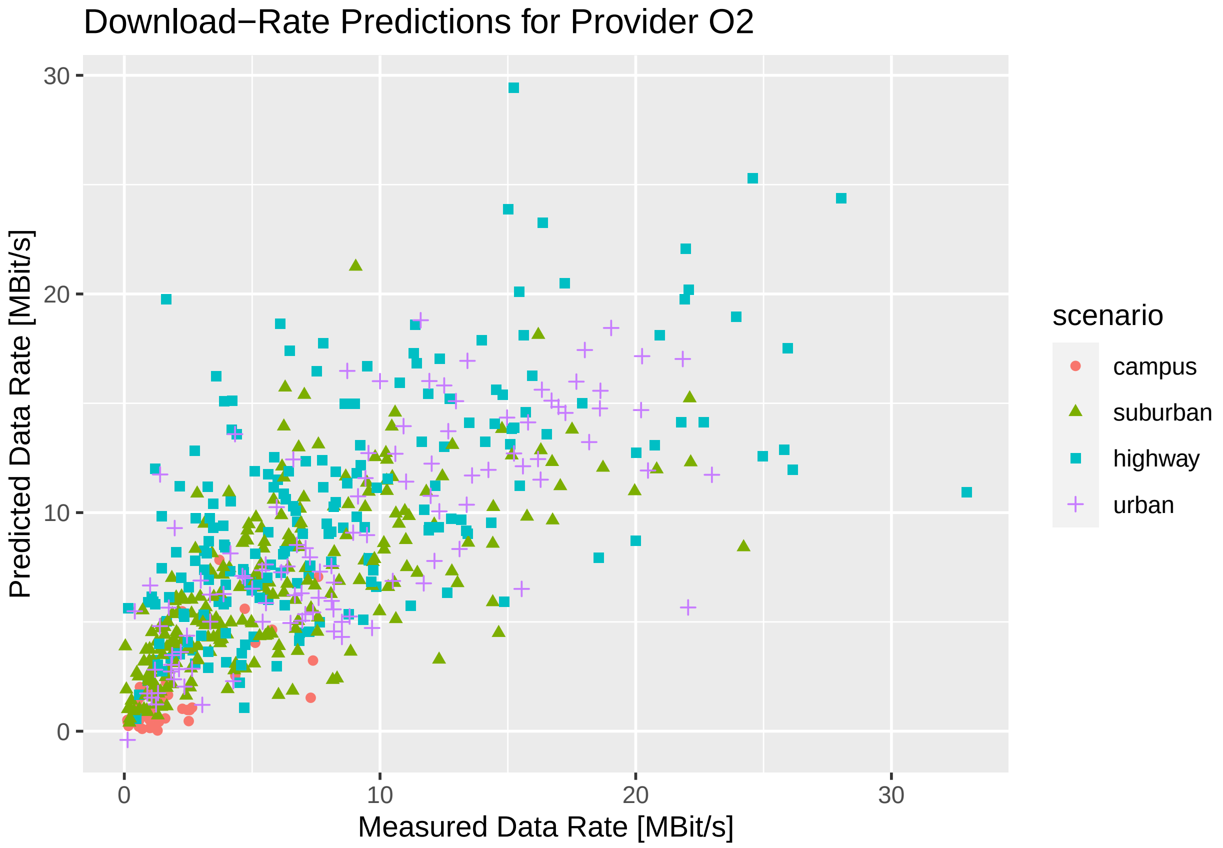
```
all(validation_results_dl$truth == validation_results_dl$throughput_mbits)
```

```
## [1] TRUE
```

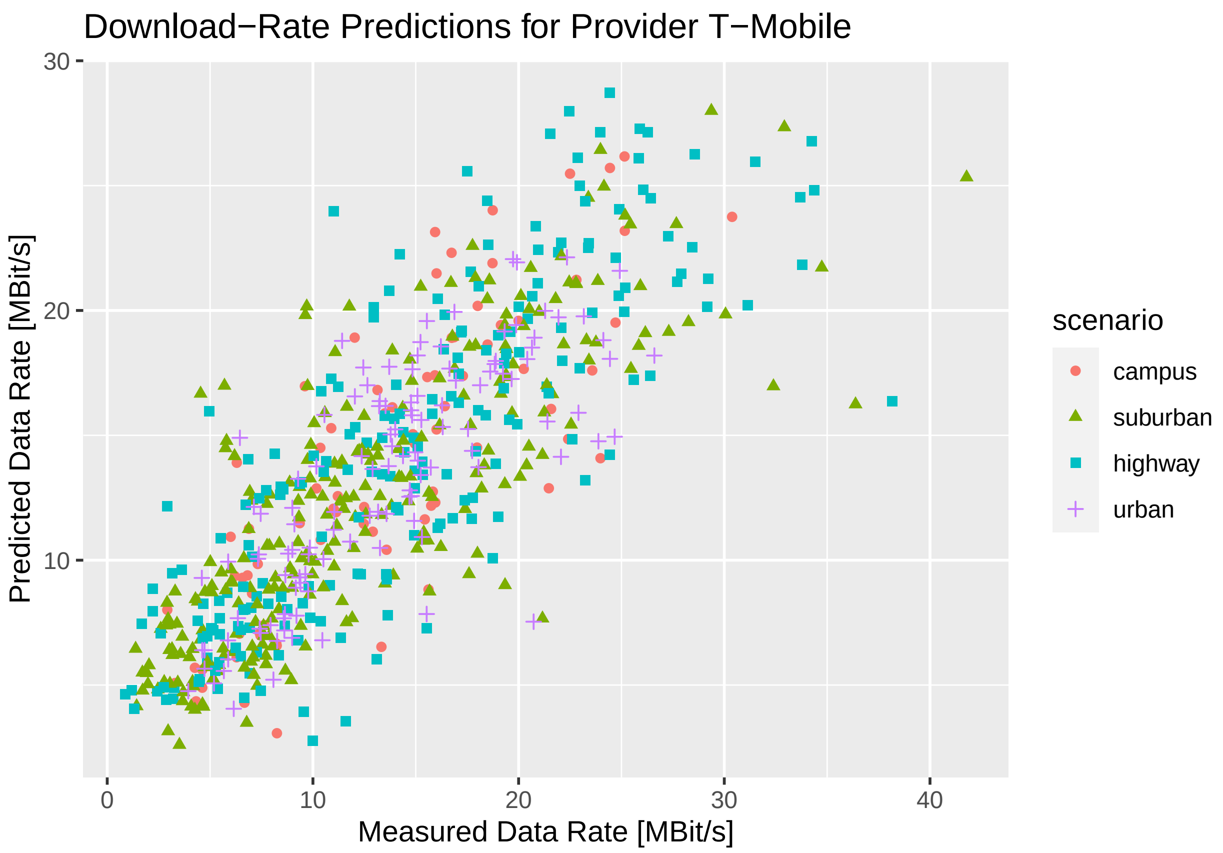
```
validation_results_dl = validation_results_dl %>%
  rename(prediction_xgboost=response) %>%
  select(-truth, -row_id, -row_name)

# write_csv(validation_results_dl, str_c(results_dir, "predictions_xgboost_dl.csv"))
```

2.5.1 Scatter Plots

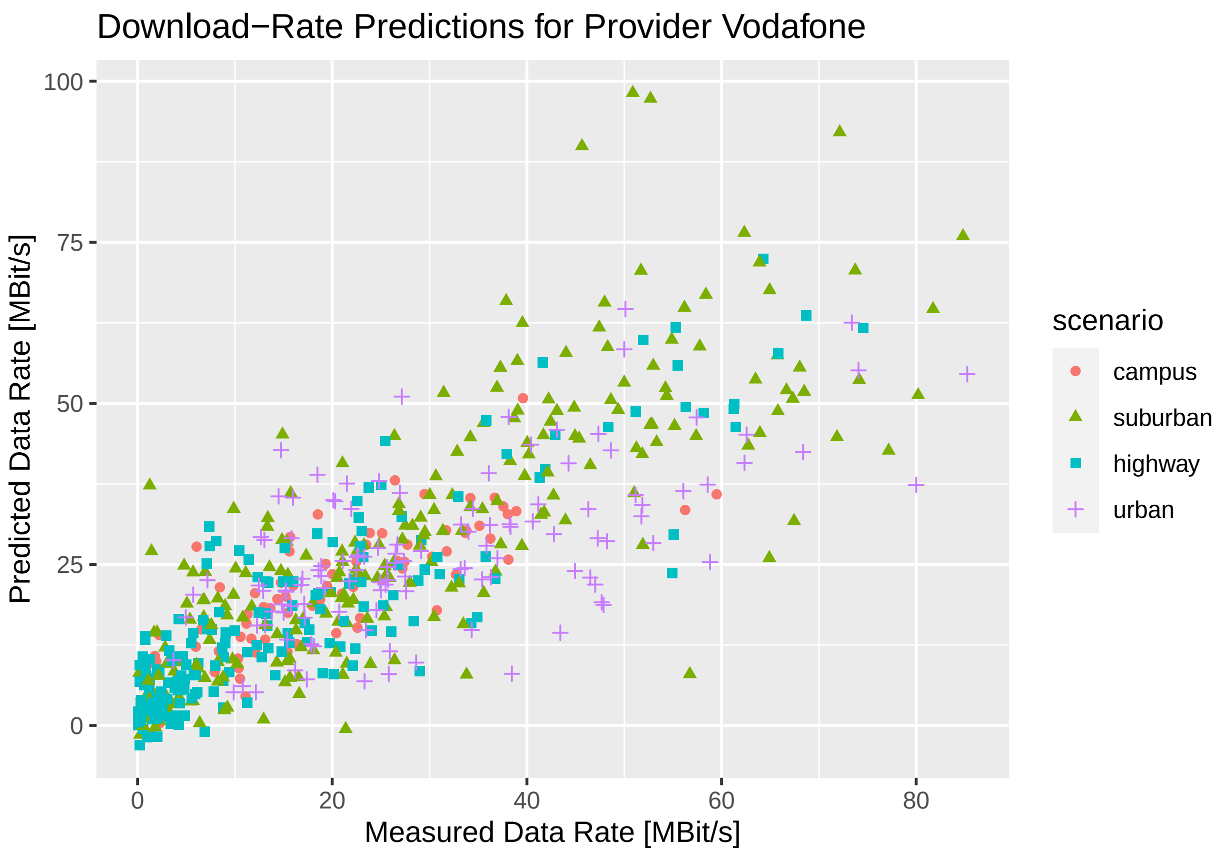
```
ggplot(filter(validation_results_dl, provider=="o2"), aes(x=throughput_mbits, y=prediction_xgboost)) +
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Data Rate [MBit/s]") +
  ylab("Predicted Data Rate [MBit/s]") +
  ggtitle("Download-Rate Predictions for Provider O2")
```



```
ggplot(filter(validation_results_dl, provider=="tmobile"), aes(x=throughput_mbits, y=prediction_xgboost
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Data Rate [MBit/s]") +
  ylab("Predicted Data Rate [MBit/s]") +
  ggtitle("Download-Rate Predictions for Provider T-Mobile")
```



```r
ggplot(filter(validation_results_dl, provider=="vodafone"), aes(x=throughput_mbits, y=prediction_xgboost
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Data Rate [MBit/s]") +
  ylab("Predicted Data Rate [MBit/s]") +
  ggtitle("Download-Rate Predictions for Provider Vodafone")
```



2.6 Feature Importance

2.6.1 XGBoost Gain

```
importance_dl_o2 = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_dl_o2$learners[[1]]$model$regr.xgboost$model$feature_names,
  model = resampling_result_dl_o2$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "o2")
```

```
## [21:05:00] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecate
```

```
importance_dl_tmobile = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_dl_tmobile$learners[[1]]$model$regr.xgboost$model$feature_names,
  model = resampling_result_dl_tmobile$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "tmobile")
```

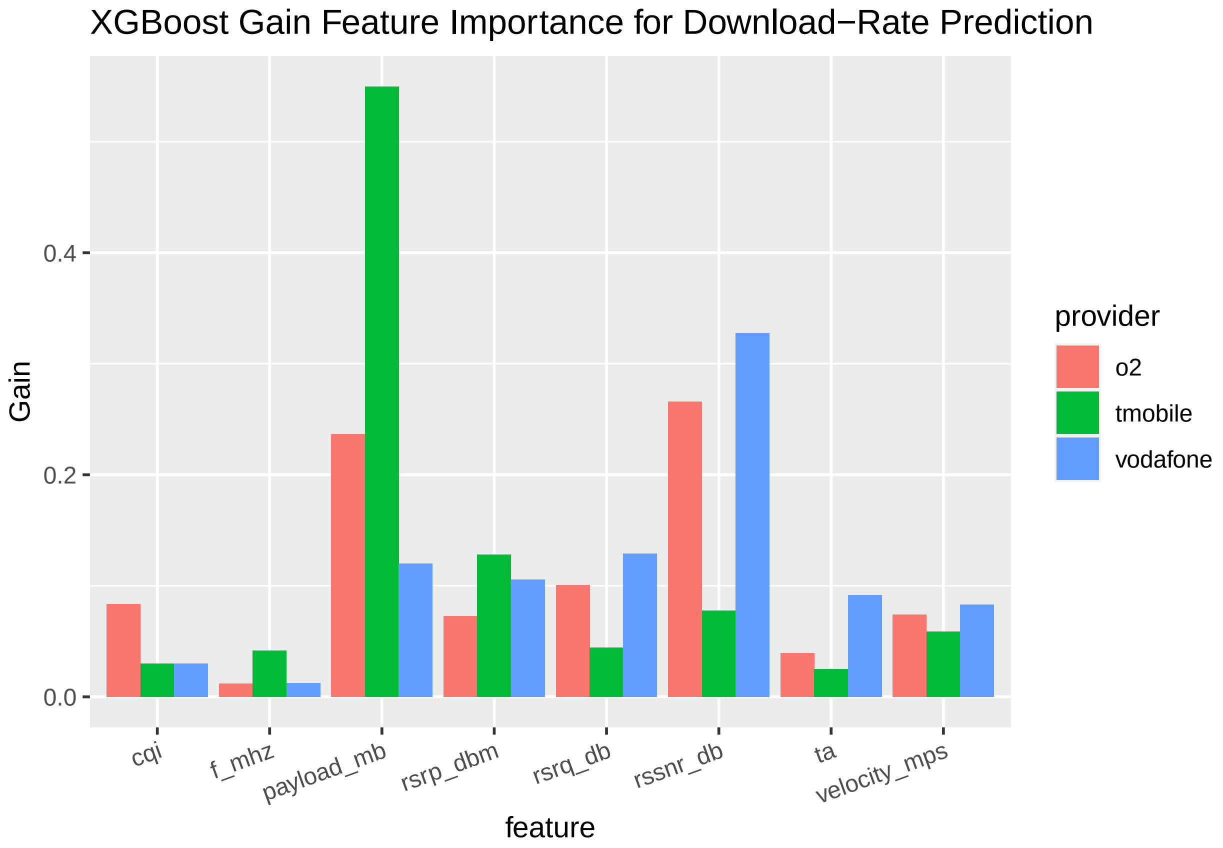
```
## [21:05:01] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecate
```

```
importance_dl_vodafone = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_dl_vodafone$learners[[1]]$model$regr.xgboost$model$feature_names,
  model = resampling_result_dl_vodafone$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "vodafone")
```

```
## [21:05:01] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecate
```

```
importance_dl = bind_rows(
  importance_dl_o2,
  importance_dl_tmobile,
  importance_dl_vodafone
) %>% filter(!str_starts(Feature, "enodeb"))
```

```
ggplot(importance_dl) +
  geom_bar(position = "dodge", aes(x = Feature, y = Gain, fill = provider), stat="identity") +
  xlab("feature") +
  ylab("Gain") +
  scale_x_discrete(guide = guide_axis(angle = 20)) +
  ggtitle("XGBoost Gain Feature Importance for Download-Rate Prediction")
```

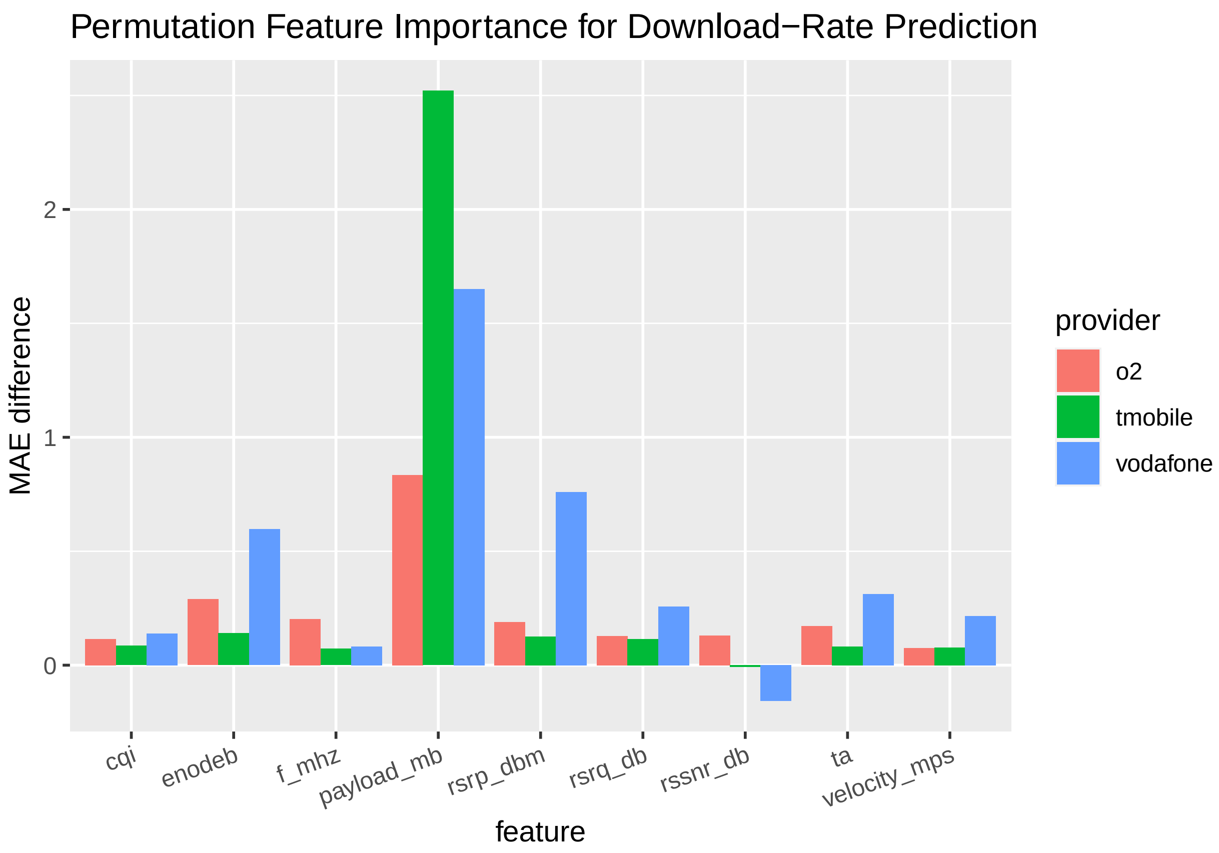


2.6.2 Permutation

```
num_permutation_sims_dl = 1

filter_permutation_o2_dl = flt("permutation",
  learner = learner_dl_o2$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_dl_o2, dataset_dl, drive_ids_train=1:7, drive_ids_test=8:10)
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc = num_permutation_sims_dl
)
filter_permutation_o2_dl$calculate(task_dl_o2)
permutation_dl_o2 = tibble(as.data.table(filter_permutation_o2_dl)) %>% mutate(provider="o2")
```

```r
filter_permutation_tmobile_dl = flt("permutation",
  learner = learner_dl_tmobile$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_dl_tmobile, dataset_dl, drive_ids_train=1:7, drive_ids_test=8:10)
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc = num_permutation_sims_dl
)
filter_permutation_tmobile_dl$calculate(task_dl_tmobile)
permutation_dl_tmobile = tibble(as.data.table(filter_permutation_tmobile_dl)) %>% mutate(provider="tmobi

filter_permutation_vodafone_dl = flt("permutation",
  learner = learner_dl_vodafone$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_dl_vodafone, dataset_dl, drive_ids_train=1:7, drive_ids_test=8:10)
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc = num_permutation_sims_dl
)
filter_permutation_vodafone_dl$calculate(task_dl_vodafone)
permutation_dl_vodafone = tibble(as.data.table(filter_permutation_vodafone_dl)) %>% mutate(provider="vod

permutation_dl = bind_rows(
  permutation_dl_o2,
  permutation_dl_tmobile,
  permutation_dl_vodafone
)
```

```r
ggplot(permutation_dl) +
  geom_bar(position = "dodge", aes(x = feature, y = score, fill = provider), stat="identity") +
  xlab("feature") +
  ylab("MAE difference") +
  scale_x_discrete(guide = guide_axis(angle = 20)) +
  ggtitle("Permutation Feature Importance for Download-Rate Prediction")
```



3 Link Lifetime Prediction

3.1 Reading the Data

```
dataset_linklifetime = read_csv(
  str_c(data_dir, "dataset_context.csv"),
  col_types = cols(
    drive_id = col_integer(),
    scenario = col_factor(),
    provider = col_factor(),
    ci = col_factor(),
    enodeb = col_factor()
  )
) %>% select(
  drive_id,
  timestamp,
  scenario,
  provider,
  velocity_mps,
  rsrp_dbm,
  rsrq_db,
  rssnr_db,
  cqi,
  ta,
  enodeb,
  rsrp_neighbor,
  rsrq_neighbor,
  link_lifetime
) %>% drop_na() %>% rowid_to_column(var="row_id_original")
```

```
dataset_linklifetime_o2 = filter(dataset_linklifetime, provider=="o2")
glimpse(dataset_linklifetime_o2)
```

```
## Rows: 22,040
## Columns: 15
## $ row_id_original <int> 1, 2, 4, 6, 8, 10, 12, 15, 18, 21, 24, 27, 30, 33, ...
## $ drive_id        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ timestamp       <dttm> 2018-12-10 09:08:46, 2018-12-10 09:08:47, 2018-12-...
## $ scenario        <fct> campus, campus, campus, campus, campus, campus, cam...
## $ provider        <fct> o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2,...
## $ velocity_mps    <dbl> 6.76, 7.65, 8.57, 10.08, 10.73, 10.93, 11.19, 11.66...
## $ rsrp_dbm        <dbl> -98, -101, -101, -94, -94, -98, -98, -94, -94, -94,...
## $ rsrq_db         <dbl> -10, -12, -12, -9, -9, -8, -8, -9, -9, -9, -9, -9, ...
## $ rssnr_db        <dbl> -1, -1, -1, 5, 5, 1, 1, -2, -2, -2, -1, -1, -3, -3,...
## $ cqi             <dbl> 9, 6, 6, 12, 12, 10, 10, 5, 5, 5, 8, 8, 6, 6, 5, 5,...
## $ ta              <dbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, ...
## $ enodeb          <fct> 54016, 54016, 54016, 54016, 54016, 54016, 54016, 54...
## $ rsrp_neighbor   <dbl> -99, -104, -104, -100, -100, -98, -98, -98, -98, -9...
## $ rsrq_neighbor   <dbl> -12, -14, -14, -17, -17, -11, -11, -11, -11, -11, -...
## $ link_lifetime   <dbl> 18.01, 17.00, 16.00, 15.00, 14.00, 13.00, 12.00, 11...
```

```
dataset_linklifetime_tmobile = filter(dataset_linklifetime, provider=="tmobile")
glimpse(dataset_linklifetime_tmobile)
```

```
## Rows: 23,741
## Columns: 15
## $ row_id_original <int> 3, 5, 7, 9, 11, 13, 16, 19, 22, 25, 28, 31, 34, 37,...
## $ drive_id        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ timestamp       <dttm> 2018-12-10 09:08:47, 2018-12-10 09:08:48, 2018-12-...
## $ scenario        <fct> campus, campus, campus, campus, campus, campus, cam...
## $ provider        <fct> tmobile, tmobile, tmobile, tmobile, tmobile, tmobil...
## $ velocity_mps    <dbl> 3.35, 3.81, 9.01, 10.84, 11.14, 11.22, 11.44, 11.75...
## $ rsrp_dbm        <dbl> -91, -91, -88, -88, -88, -86, -86, -88, -88, -85, -...
## $ rsrq_db         <dbl> -6, -6, -6, -6, -6, -6, -6, -7, -7, -5, -5, -5, -5,...
## $ rssnr_db        <dbl> 12, 12, 18, 18, 18, 20, 20, 19, 19, 22, 22, 18, 18,...
## $ cqi             <dbl> 11, 11, 15, 15, 15, 12, 12, 15, 15, 10, 10, 10, 10,...
## $ ta              <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, ...
## $ enodeb          <fct> 103068, 103068, 103068, 103068, 103068, 103068, 103...
## $ rsrp_neighbor   <dbl> -Inf, -Inf, -Inf, -Inf, -Inf, -Inf, -Inf, -Inf, -In...
## $ rsrq_neighbor   <dbl> -Inf, -Inf, -Inf, -Inf, -Inf, -Inf, -Inf, -Inf, -In...
## $ link_lifetime   <dbl> 19.01, 18.00, 16.99, 16.00, 15.00, 14.01, 13.01, 12...
```

```
dataset_linklifetime_vodafone = filter(dataset_linklifetime, provider=="vodafone")
glimpse(dataset_linklifetime_vodafone)
```

```
## Rows: 22,553
## Columns: 15
## $ row_id_original <int> 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 57, 60,...
## $ drive_id        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ timestamp       <dttm> 2018-12-10 09:08:52, 2018-12-10 09:08:53, 2018-12-...
## $ scenario        <fct> campus, campus, campus, campus, campus, campus, cam...
## $ provider        <fct> vodafone, vodafone, vodafone, vodafone, vodafone, v...
## $ velocity_mps    <dbl> 3.45, 3.45, 3.88, 9.69, 10.89, 11.01, 11.33, 11.63,...
## $ rsrp_dbm        <dbl> -109, -114, -114, -120, -120, -117, -117, -118, -11...
## $ rsrq_db         <dbl> -9, -10, -10, -18, -18, -11, -11, -11, -11, -15, -1...
## $ rssnr_db        <dbl> 3, 7, 7, -4, -4, -6, -6, -16, -16, -8, -8, -5, -5, ...
## $ cqi             <dbl> 10, 10, 10, 4, 4, 7, 7, 3, 3, 4, 4, 6, 6, 6, 6, 6, ...
## $ ta              <dbl> 24, 24, 24, 63, 63, 63, 63, 63, 63, 63, 63, 17, 17,...
## $ enodeb          <fct> 52316, 52316, 52316, 51044, 51044, 51044, 51044, 51...
## $ rsrp_neighbor   <dbl> -Inf, -115, -115, -111, -111, -111, -111, -119, -11...
## $ rsrq_neighbor   <dbl> -Inf, -12, -12, -13, -13, -13, -13, -14, -14, -13, ...
## $ link_lifetime   <dbl> 3.01, 2.00, 1.00, 13.00, 12.00, 11.00, 10.00, 9.00,...
```

3.2 Create the Prediction Tasks for Each Provider

```
task_linklifetime_o2 = make_task(dataset_linklifetime_o2, "task_linklifetime_o2", target = "link_lifeti
task_linklifetime_o2
```

```
## <TaskRegr:task_linklifetime_o2> (22040 x 10)
## * Target: link_lifetime
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, rsrp_dbm, rsrp_neighbor, rsrq_db, rsrq_neighbor,
##     rssnr_db, ta, velocity_mps
##   - fct (1): enodeb
```

```
task_linklifetime_tmobile = make_task(dataset_linklifetime_tmobile, "task_linklifetime_tmobile", target
task_linklifetime_tmobile
```

```
## <TaskRegr:task_linklifetime_tmobile> (23741 x 10)
## * Target: link_lifetime
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, rsrp_dbm, rsrp_neighbor, rsrq_db, rsrq_neighbor,
##     rssnr_db, ta, velocity_mps
##   - fct (1): enodeb
```

```
task_linklifetime_vodafone = make_task(dataset_linklifetime_vodafone, "task_linklifetime_vodafone", targ
task_linklifetime_vodafone
```

```
## <TaskRegr:task_linklifetime_vodafone> (22553 x 10)
## * Target: link_lifetime
## * Properties: -
## * Features (9):
##   - dbl (8): cqi, rsrp_dbm, rsrp_neighbor, rsrq_db, rsrq_neighbor,
##     rssnr_db, ta, velocity_mps
##   - fct (1): enodeb
```

3.3 Parameter Tuning

```
tuning_result_linklifetime_o2 = get_tuning_result(task_linklifetime_o2, dataset_linklifetime, grid_reso

tuning_result_linklifetime_tmobile = get_tuning_result(task_linklifetime_tmobile, dataset_linklifetime,

tuning_result_linklifetime_vodafone = get_tuning_result(task_linklifetime_vodafone, dataset_linklifetim
tuning_result_linklifetime = bind_rows(
  tibble(tuning_result_linklifetime_o2$result) %>% mutate(provider="o2"),
  tibble(tuning_result_linklifetime_tmobile$result) %>% mutate(provider="tmobile"),
  tibble(tuning_result_linklifetime_vodafone$result) %>% mutate(provider="vodafone"),
) %>% select("provider", "regr.xgboost.nrounds", "regr.xgboost.eta", "regr.xgboost.gamma", "regr.xgboos

knitr::kable(tuning_result_linklifetime)
```

provider	regr.xgboost.nrounds	regr.xgboost.eta	regr.xgboost.gamma	regr.xgboost.lambda
o2	669	0.2184211	4.210526	8.421053
tmobile	669	0.0621053	5.263158	10.000000
vodafone	431	0.0621053	2.631579	9.473684

3.4 Create Learners with Tuned Hyperparameters

```
learner_linklifetime_o2 = make_learner(
  nrounds = tuning_result_linklifetime_o2$result$regr.xgboost.nrounds,
  eta = tuning_result_linklifetime_o2$result$regr.xgboost.eta,
  gamma = tuning_result_linklifetime_o2$result$regr.xgboost.gamma,
  lambda = tuning_result_linklifetime_o2$result$regr.xgboost.lambda
)

learner_linklifetime_tmobile = make_learner(
  nrounds = tuning_result_linklifetime_tmobile$result$regr.xgboost.nrounds,
  eta = tuning_result_linklifetime_tmobile$result$regr.xgboost.eta,
  gamma = tuning_result_linklifetime_tmobile$result$regr.xgboost.gamma,
  lambda = tuning_result_linklifetime_tmobile$result$regr.xgboost.lambda
)

learner_linklifetime_vodafone = make_learner(
  nrounds = tuning_result_linklifetime_vodafone$result$regr.xgboost.nrounds,
  eta = tuning_result_linklifetime_vodafone$result$regr.xgboost.eta,
  gamma = tuning_result_linklifetime_vodafone$result$regr.xgboost.gamma,
  lambda = tuning_result_linklifetime_vodafone$result$regr.xgboost.lambda
)
```

3.5 Validation Results

```
resampling_result_linklifetime_o2 = resample(
  task = task_linklifetime_o2,
  learner = learner_linklifetime_o2,
  resampling = make_outer_resampling(task_linklifetime_o2, dataset_linklifetime, drive_ids_train=1:7, d
  store_models = TRUE
)

resampling_result_linklifetime_tmobile = resample(
  task = task_linklifetime_tmobile,
  learner = learner_linklifetime_tmobile,
  resampling = make_outer_resampling(task_linklifetime_tmobile, dataset_linklifetime, drive_ids_train=1
  store_models = TRUE
)
```

```r
resampling_result_linklifetime_vodafone = resample(
  task = task_linklifetime_vodafone,
  learner = learner_linklifetime_vodafone,
  resampling = make_outer_resampling(task_linklifetime_vodafone, dataset_linklifetime, drive_ids_train=
  store_models = TRUE
)
```

```r
predictions_linklifetime_o2 = as.data.table(resampling_result_linklifetime_o2$prediction())
glimpse(tibble(predictions_linklifetime_o2))
```

```
## Rows: 6,610
## Columns: 3
## $ row_id   <int> 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735, 1736...
## $ truth    <dbl> 59.01, 58.01, 57.00, 56.00, 55.00, 54.01, 53.01, 52.01, 51...
## $ response <dbl> 62.65211, 62.65211, 76.24164, 76.24164, 76.24164, 67.36938...
```

```r
predictions_linklifetime_tmobile = as.data.table(resampling_result_linklifetime_tmobile$prediction())
predictions_linklifetime_vodafone = as.data.table(resampling_result_linklifetime_vodafone$prediction())
```

```r
validation_results_linklifetime = bind_rows(
  tibble(predictions_linklifetime_o2) %>%
    inner_join(tibble(task_linklifetime_o2$row_names), by="row_id") %>%
    inner_join(dataset_linklifetime, by=c("row_name"="row_id_original")),
  tibble(predictions_linklifetime_tmobile) %>%
    inner_join(tibble(task_linklifetime_tmobile$row_names), by="row_id") %>%
    inner_join(dataset_linklifetime, by=c("row_name"="row_id_original")),
  tibble(predictions_linklifetime_vodafone) %>%
    inner_join(tibble(task_linklifetime_vodafone$row_names), by="row_id") %>%
    inner_join(dataset_linklifetime, by=c("row_name"="row_id_original"))
)
glimpse(validation_results_linklifetime)
```

```
## Rows: 20,207
## Columns: 18
## $ row_id        <int> 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735,...
## $ truth         <dbl> 59.01, 58.01, 57.00, 56.00, 55.00, 54.01, 53.01, 52.0...
## $ response      <dbl> 62.65211, 62.65211, 76.24164, 76.24164, 76.24164, 67....
## $ row_name      <int> 6247, 6250, 6253, 6256, 6259, 6262, 6265, 6268, 6271,...
## $ drive_id      <int> 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8,...
## $ timestamp     <dttm> 2018-12-11 09:04:04, 2018-12-11 09:04:05, 2018-12-11...
## $ scenario      <fct> campus, campus, campus, campus, campus, campus, campu...
## $ provider      <fct> o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o2, o...
## $ velocity_mps  <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00,...
## $ rsrp_dbm      <dbl> -88, -88, -89, -89, -89, -89, -89, -86, -86, -91, -91...
## $ rsrq_db       <dbl> -11, -11, -8, -8, -8, -9, -9, -7, -7, -9, -9, -13, -1...
## $ rssnr_db      <dbl> 7, 7, 12, 12, 12, 13, 13, 11, 11, 6, 6, 4, 4, 4, 4, 3...
## $ cqi           <dbl> 2, 2, 8, 8, 8, 11, 11, 9, 9, 7, 7, 6, 6, 6, 6, 6, 6, ...
## $ ta            <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,...
## $ enodeb        <fct> 52410, 52410, 52410, 52410, 52410, 52410, 52410, 5241...
## $ rsrp_neighbor <dbl> -Inf, -Inf, -95, -95, -95, -95, -95, -95, -95, -95, -...
## $ rsrq_neighbor <dbl> -Inf, -Inf, -13, -13, -13, -13, -13, -13, -13, -13, -...
## $ link_lifetime <dbl> 59.01, 58.01, 57.00, 56.00, 55.00, 54.01, 53.01, 52.0...
```

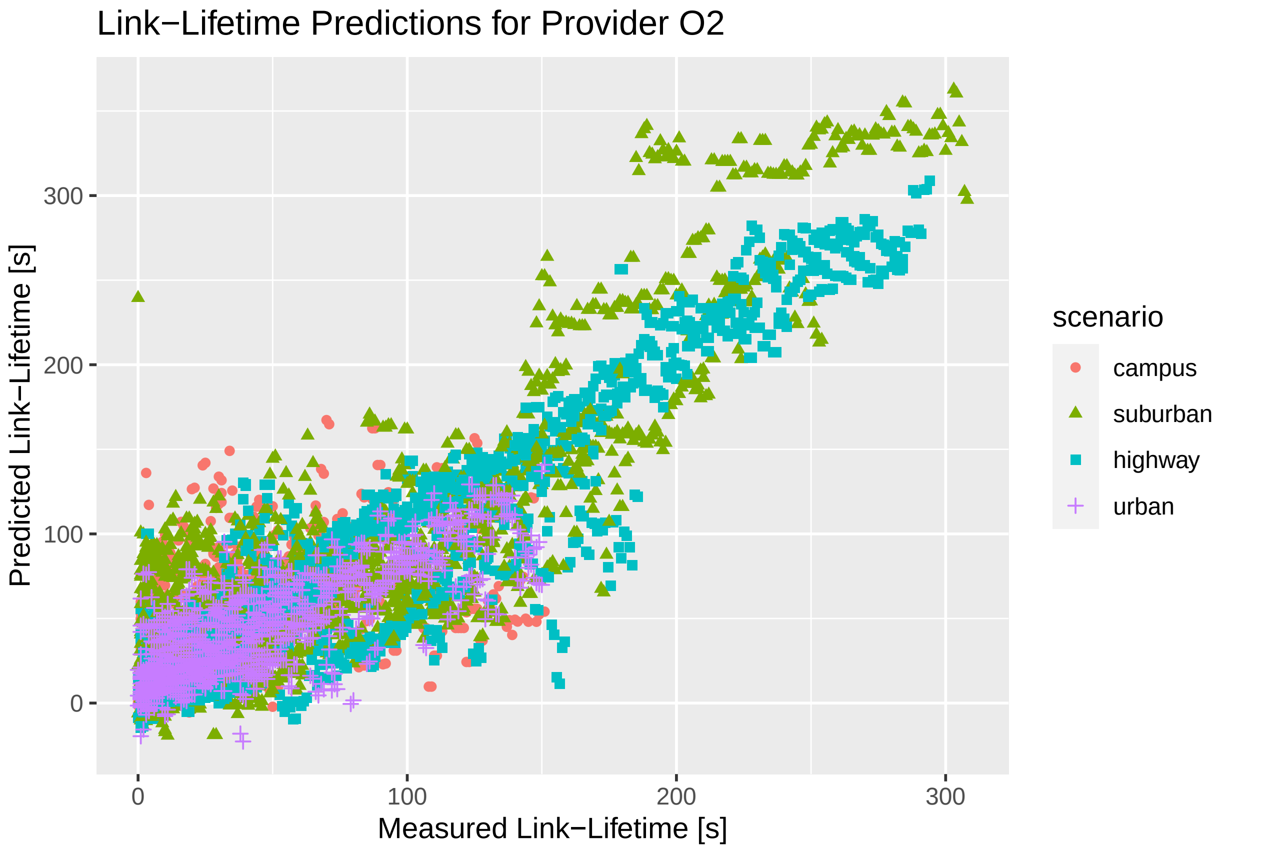
```
all(validation_results_linklifetime$truth == validation_results_linklifetime$link_lifetime)
```

```
## [1] TRUE
```

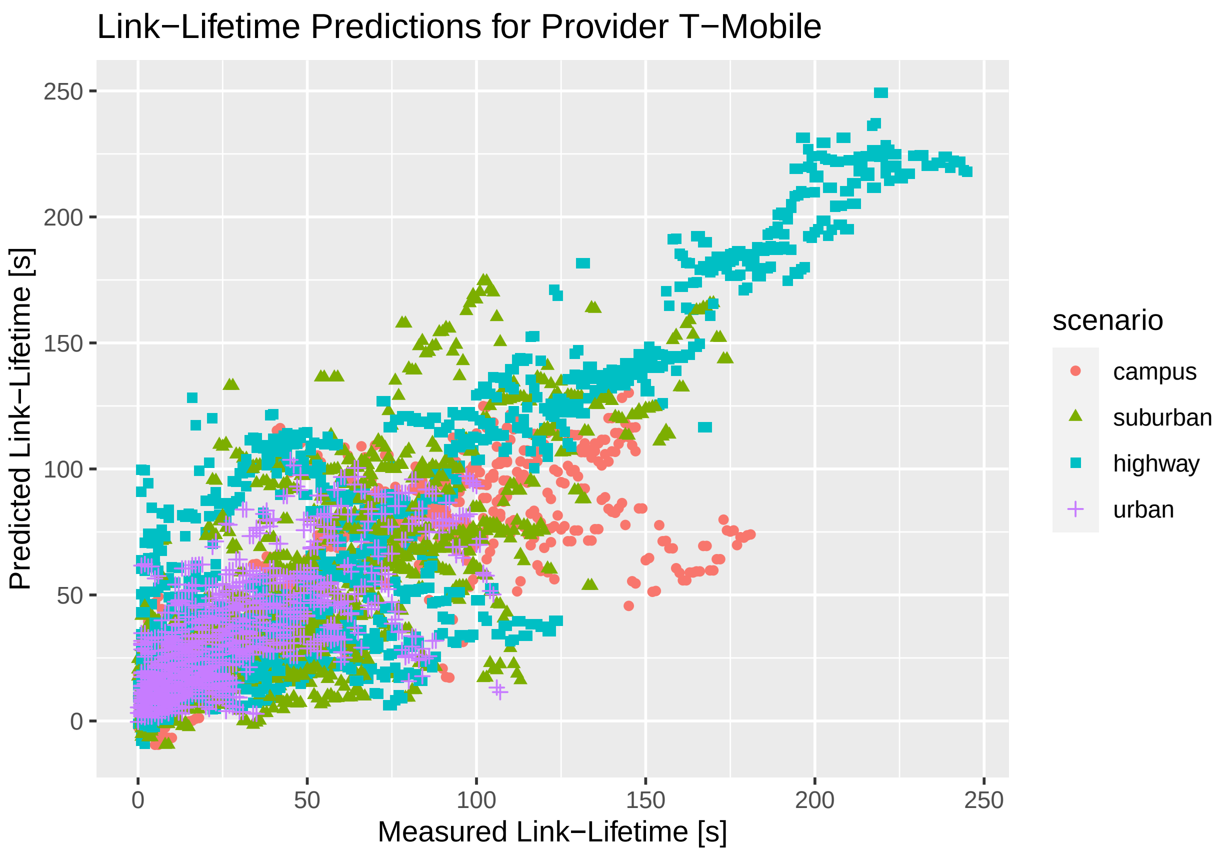
```
validation_results_linklifetime = validation_results_linklifetime %>%
  rename(prediction_xgboost=response) %>%
  select(-truth, -row_id, -row_name)

# write_csv(validation_results_linklifetime, str_c(results_dir, "predictions_xgboost_linklifetime.csv")
```

3.5.1 Scatter Plots

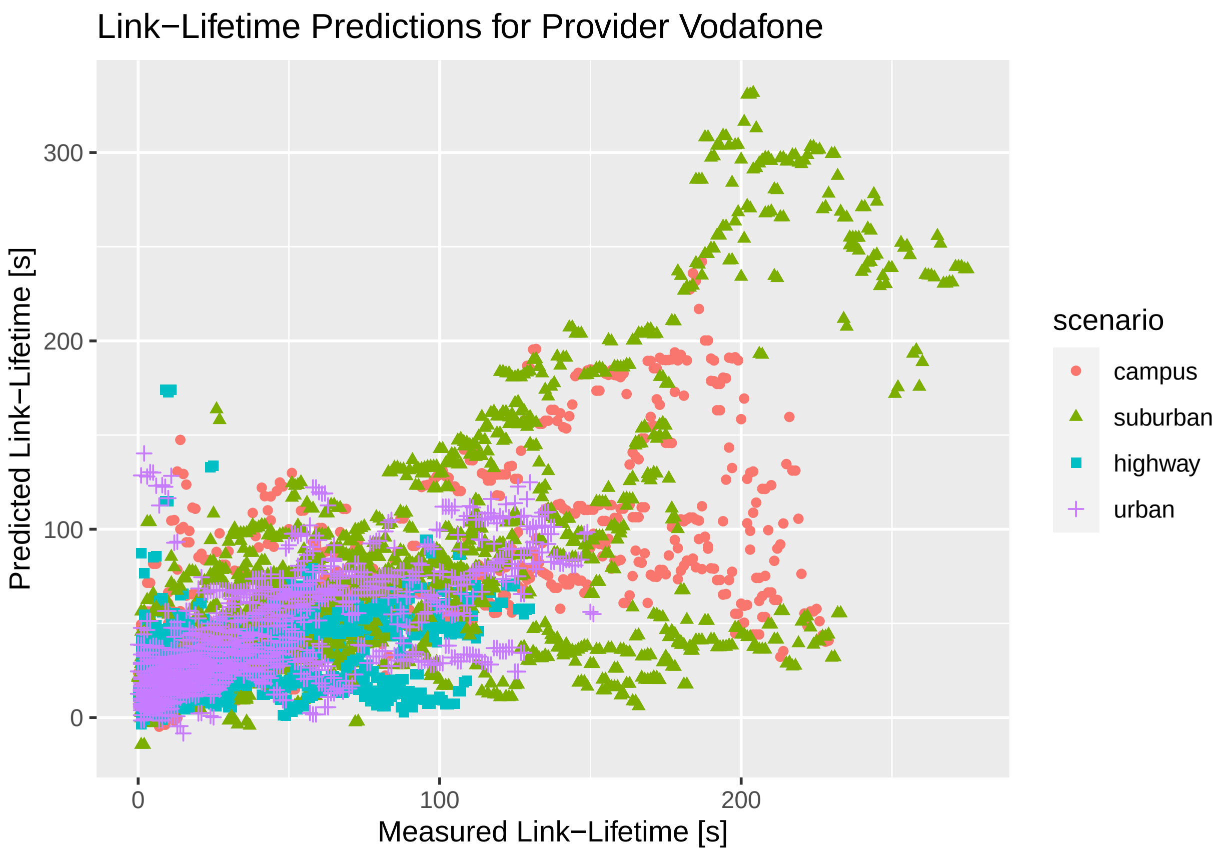
```
ggplot(filter(validation_results_linklifetime, provider=="o2"), aes(x=link_lifetime, y=prediction_xgboo
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Link-Lifetime [s]") +
  ylab("Predicted Link-Lifetime [s]") +
  ggtitle("Link-Lifetime Predictions for Provider O2")
```



```
ggplot(filter(validation_results_linklifetime, provider=="tmobile"), aes(x=link_lifetime, y=prediction_
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Link-Lifetime [s]") +
  ylab("Predicted Link-Lifetime [s]") +
  ggtitle("Link-Lifetime Predictions for Provider T-Mobile")
```



```r
ggplot(filter(validation_results_linklifetime, provider=="vodafone"), aes(x=link_lifetime, y=prediction
  geom_point(aes(color=scenario, shape=scenario)) +
  xlab("Measured Link-Lifetime [s]") +
  ylab("Predicted Link-Lifetime [s]") +
  ggtitle("Link-Lifetime Predictions for Provider Vodafone")
```



3.6 Feature Importance

3.6.1 XGBoost Gain

```
importance_linklifetime_o2 = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_linklifetime_o2$learners[[1]]$model$regr.xgboost$model$feature_name
  model = resampling_result_linklifetime_o2$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "o2")
```

```
## [21:12:21] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecate
```

```
importance_linklifetime_tmobile = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_linklifetime_tmobile$learners[[1]]$model$regr.xgboost$model$feature_
  model = resampling_result_linklifetime_tmobile$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "tmobile")
```

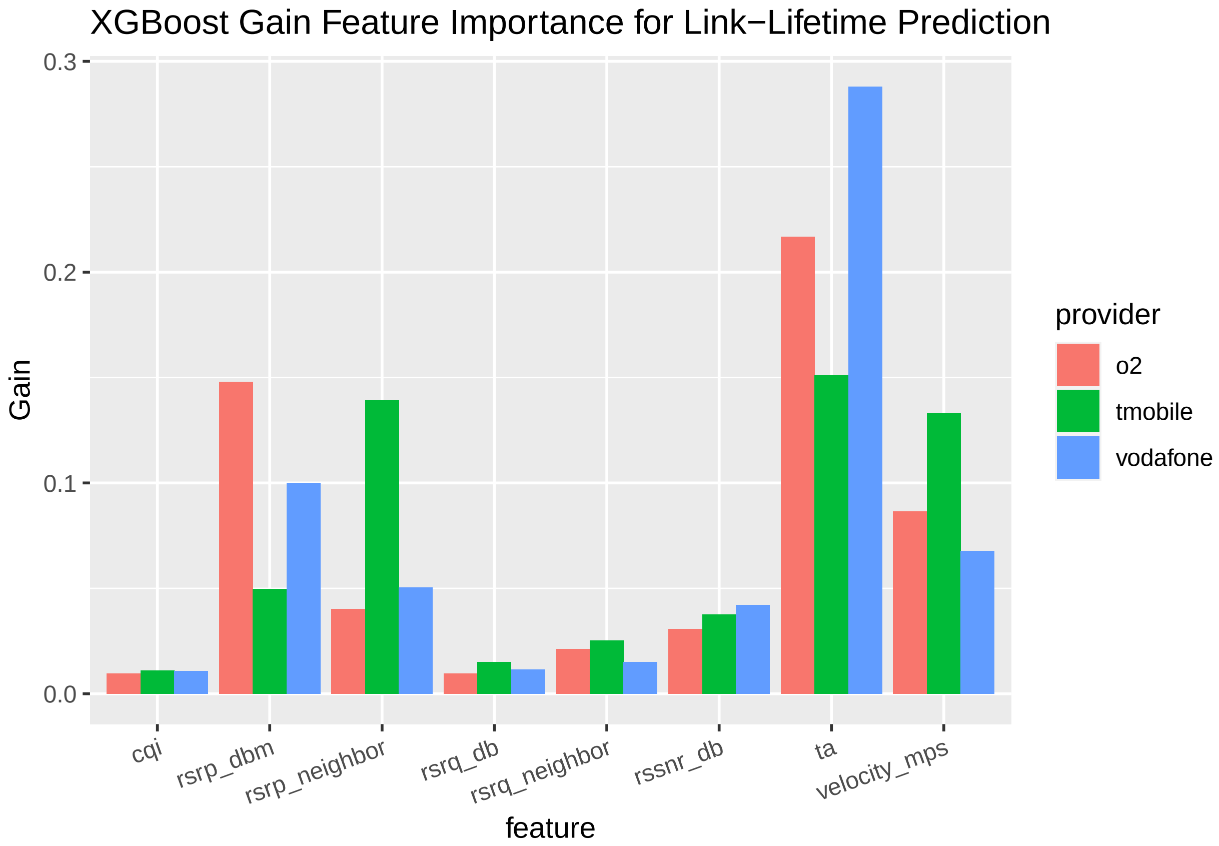
```
## [21:12:22] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecate
```

```
importance_linklifetime_vodafone = tibble(xgboost::xgb.importance(
  feature_names = resampling_result_linklifetime_vodafone$learners[[1]]$model$regr.xgboost$model$featur
  model = resampling_result_linklifetime_vodafone$learners[[1]]$model$regr.xgboost$model
)) %>% mutate(provider = "vodafone")
```

```
## [21:12:23] WARNING: amalgamation/../src/objective/regression_obj.cu:174: reg:linear is now deprecate
```

```
importance_linklifetime = bind_rows(
  importance_linklifetime_o2,
  importance_linklifetime_tmobile,
  importance_linklifetime_vodafone
) %>% filter(!str_starts(Feature, "enodeb"))
```

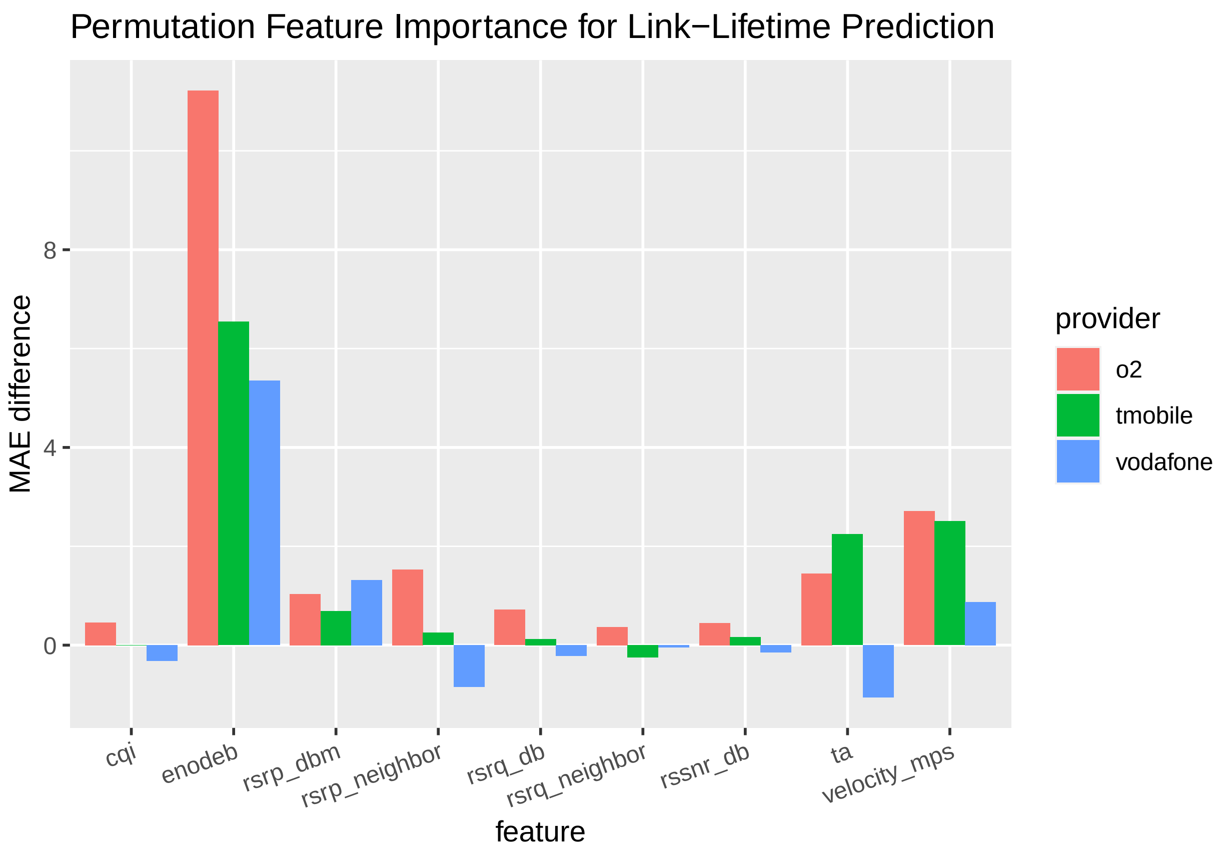
```
ggplot(importance_linklifetime) +
  geom_bar(position = "dodge", aes(x = Feature, y = Gain, fill = provider), stat="identity") +
  xlab("feature") +
  ylab("Gain") +
  scale_x_discrete(guide = guide_axis(angle = 20)) +
  ggtitle("XGBoost Gain Feature Importance for Link-Lifetime Prediction")
```



3.6.2 Permutation

```
num_permutation_sims_linklifetime = 1

filter_permutation_o2_linklifetime = flt("permutation",
  learner = learner_linklifetime_o2$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_linklifetime_o2, dataset_linklifetime, drive_ids_train=1:7, drive_ids_te
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc = num_permutation_sims_linklifetime
)
filter_permutation_o2_linklifetime$calculate(task_linklifetime_o2)
permutation_linklifetime_o2 = tibble(as.data.table(filter_permutation_o2_linklifetime)) %>% mutate(provi
filter_permutation_tmobile_linklifetime = flt("permutation",
  learner = learner_linklifetime_tmobile$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_linklifetime_tmobile, dataset_linklifetime, drive_ids_train=1:7, drive_i
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc = num_permutation_sims_linklifetime
)
filter_permutation_tmobile_linklifetime$calculate(task_linklifetime_tmobile)
permutation_linklifetime_tmobile = tibble(as.data.table(filter_permutation_tmobile_linklifetime)) %>% mu

filter_permutation_vodafone_linklifetime = flt("permutation",
  learner = learner_linklifetime_vodafone$clone(),
  resampling = uninstantiate_resampling(
    make_outer_resampling(task_linklifetime_vodafone, dataset_linklifetime, drive_ids_train=1:7, drive_
  ),
  measure = msr("regr.mae"),
  standardize = TRUE,
  nmc = num_permutation_sims_linklifetime
)
filter_permutation_vodafone_linklifetime$calculate(task_linklifetime_vodafone)
permutation_linklifetime_vodafone = tibble(as.data.table(filter_permutation_vodafone_linklifetime)) %>%

permutation_linklifetime = bind_rows(
  permutation_linklifetime_o2,
  permutation_linklifetime_tmobile,
  permutation_linklifetime_vodafone
)
```

```
ggplot(permutation_linklifetime) +
  geom_bar(position = "dodge", aes(x = feature, y = score, fill = provider), stat="identity") +
  xlab("feature") +
  ylab("MAE difference") +
  scale_x_discrete(guide = guide_axis(angle = 20)) +
  ggtitle("Permutation Feature Importance for Link-Lifetime Prediction")
```


Permutation Feature Importance for Link−Lifetime Prediction
MAE difference
8
4
0
cqi
enodeb
rsrp_dbm
rsrp_neighbor
rsrq_db
rsrq_neighbor
rssnr_db
ta
velocity_mps
feature
provider
o2
tmobile
vodafone