

МИНИСТЕРСТВО ВЫСШЕГО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**



Кафедра Вычислительной техники

Лабораторная работа № 1-2

по дисциплине «Интеллектуальный анализ данных и машинное обучение»

«Использование платформы H2O

для интеллектуального анализа данных и машинного обучения.

Обучение, тестирование и улучшение моделей.

Загрузка моделей в базу знаний и получение

их уникальных идентификаторов.»

Вариант: 5

Преподаватель: Яковина И.Н.

Группа: АММ-19

Студенты: Высоцкий Е.В.

Новосибирск

2021 г.

Содержание

1 Цель работы	4
2 Задание	4
3 Итоговая таблица моделей	5
4 Ход работы	6
4.1 Загрузка исходных данных в систему H2O	6
4.2 Парсинг загруженных в систему данных	6
4.3 Эксперимент № 1. Метод Y1 – Gradient Boosting Machine (GBM) для полной модели «05_01_Bk_MTarg_m1_F».....	7
4.3.1 Разбиение данных для обучения и проверки модели.....	7
4.3.2 Построение полной модели «05_01_Bk_m1_MTarg_F».....	7
4.4 Эксперимент № 2. Метод Y1 – Gradient Boosting Machine (GBM) для усечённой модели «05_02_Bk_m1_MTarg_S»	8
4.4.1 Разбиение данных для обучения и проверки модели	8
4.4.2 Построение усечённой модели «05_02_Bk_m1_MTarg_S»	8
4.5 Эксперимент № 3. Метод Y1 – Generalized Linear Model (GLM) - binomial: Logistic Regression для полной модели «05_03_Bk_m8_MTarg_F».....	9
4.5.1 Разбиение данных для обучения и проверки модели	9
4.5.2 Построение полной модели «05_03_Bk_m8_MTarg_F»	9
4.6 Эксперимент № 4. Метод Y1 – Generalized Linear Model (GLM) - binomial: Logistic Regression для усечённой модели «05_04_Bk_m8_MTarg_S».....	10
4.6.1 Разбиение данных для обучения и проверки модели	10
4.6.2 Построение усечённой модели «05_04_Bk_m8_MTarg_S»	10
4.7 Эксперимент № 5. Метод Y2 – Gradient Boosting Machine (GBM) для полной модели «05_05_Mk_m1_Targ3m_F».....	11
4.7.1 Разбиение данных для обучения и проверки модели.....	11
4.7.2 Построение полной модели «05_05_Mk_m1_Targ3m_F»	11
4.8 Эксперимент № 6. Метод Y2 – Gradient Boosting Machine (GBM) для усечённой модели «05_06_Mk_m1_Targ3m_S»	12
4.8.1 Разбиение данных для обучения и проверки модели	12
4.8.2 Построение усечённой модели «05_06_Mk_m1_Targ3m_S»	12
4.9 Эксперимент № 7. Метод Y2 – Generalized Linear Model (GLM) - multinomial: Logistic Regression для полной модели «05_07_Mk_m9_Targ3m_F»	13
4.9.1 Разбиение данных для обучения и проверки модели	13
4.9.2 Построение полной модели «05_07_Mk_m9_Targ3m_F»	13
4.10 Эксперимент № 8. Метод Y2 – Generalized Linear Model (GLM) - multinomial: Logistic Regression для усечённой модели «05_08_Mk_m9_Targ3m_S»	14
4.10.1 Разбиение данных для обучения и проверки модели	14
4.10.2 Построение усечённой модели «05_08_Mk_m9_Targ3m_S»	14
4.11 Эксперимент № 9. Метод Y3 – Generalized Linear Model (GLM) - gaussian: Gaussian regression для полной модели «05_09_Rr_m4_Targ3r_F»	15
4.11.1 Разбиение данных для обучения и проверки модели	15
4.11.2 Построение полной модели «05_09_Rr_m4_Targ3r_F»	15

4.12 Эксперимент № 10. Метод Y3 – Generalized Linear Model (GLM) - gaussian:	
Gaussian regression для усечённой модели «05_10_Rr_m4_Targ3r_S»	16
4.12.1 Разбиение данных для обучения и проверки модели	16
4.12.2 Построение усечённой модели «05_10_Rr_m4_Targ3r_S»	16
4.13 Эксперимент № 11. Метод Y3 – Generalized Linear Model (GLM) - gamma:	
Gamma Models для полной модели «05_11_Rr_m7_Targ3r_F»	17
4.13.1 Разбиение данных для обучения и проверки модели	17
4.13.2 Построение полной модели «05_11_Rr_m7_Targ3r_F».....	17
4.14 Эксперимент № 12. Метод Y3 – Generalized Linear Model (GLM) - gamma:	
Gamma Models для усечённой модели «05_10_Rr_m4_Targ3r_S».....	18
4.14.1 Разбиение данных для обучения и проверки модели.....	18
4.14.2 Построение усечённой модели «05_10_Rr_m4_Targ3r_S».....	18
4.15 Анализ информации о созданных моделях.....	19
4.15.1 Важность признаков в графическом виде.....	19
4.15.2 Важность признаков в числовом виде.....	21
4.15.3 Матрицы ошибок.....	22
4.15.4 ROC – кривая.....	24
4.15.5 Метрики модели.....	26
4.16 Загрузка моделей в базу знаний и получение их уникальных идентификаторов	
UUID.....	30
Выводы по работе	31
Список литературы.....	33

1 Цель работы

Изучить программные средства и методы платформы H2O для создания, обучения и использования моделей регрессии, бинарной и множественной классификации исходных данных, согласно варианта – на примере моделирования диагностики ректального рака. Изучить методы загрузки моделей в базу знаний и опробовать один из них.

2 Задание

2.1 Загрузить исходные данные для анализа в сеансе работы платформы H2O.

2.2 Проверить визуальным осмотром в исходном файле тип разделителя ячеек, десятичную точку или запятую, полноту данных, исключить неполные строки значений во избежание появления несуществующих зависимостей. Исключить из обработки столбцы с «*id*». Проверить типы данных (*Numeric* – для регрессии, *Enum* – для классификации). Настроить соответствующие параметры в системе.

2.3 Установить один, не исключенный целевой признак в «*responce_column*».

2.4 Создать и обучить модели в соответствии с вариантом задания – табл. 1, руководствуясь обозначением методов в табл. 3, зафиксировать результаты в таблицу, используя формат имён файлов моделей, указанный в табл. 2.

2.5 Сформировать комплекс характеристик моделей и загрузить их в базу знаний.

2.6 Проверить работу моделей полных и усеченных до 10...15 значимых признаков.

2.7 Зафиксировать результаты в таблицу, применив специальные кодировки параметров.

2.8 Сделать выводы по работе.

Таблица 1. Параметры варианта исходных данных.

Вариант	Группа	Фамилия	Бригада	Файл	Y1	Метод Y1	Y2	Метод Y2	Y3	Метод Y3	Имя набора данных
05	АММ-19	Высоцкий		KPP	MTarg	m1, m8	Targ3m	m1, m9	Targ3r	m4, m7	1CRR-v2.csv

Таблица 2. Формат имени модели/файла модели.

Номер варианта	Номер эксперимента	Задача (Bk/Mk/Rr) Bk - Классиф. бинарн. Mk - Классиф. множ. Rr- Регрессия	Метод (m1/m2/.../m9)	Целевая переменная (MTarg, Targ1m, ...)	Набор признаков (F – Полный, S – Усеченный)	Примечания
xx	xx	xx	xx	xxxx	x	кол-во символов
01	01	Bk	m1	MTarg	F	пример
01_01_Bk_m1_MTarg_F						

Таблица 3. Методы анализа данных.

Код	Классиф. бинарная (Bk)	Классиф. множ. (Mk)	Регрессия (Rr)	Имена методов
m1	Bk	Mk	Rr	Gradient Boosting Machine (GBM)
m2	Bk	Mk	Rr	Distributed Random Forest (DRF)
m3	Bk	Mk	-	Naïve Bayes
m4	-	-	Rr	Generalized Linear Model (GLM) - gaussian: Gaussian regression
m5	-	-	Rr	Generalized Linear Model (GLM) - quasibinomial: Pseudo-Logistic Regression (Quasibinomial Family)
m6	-	-	Rr	Generalized Linear Model (GLM) - Poisson regression
m7	-	-	Rr	Generalized Linear Model (GLM) - gamma: Gamma Models
m8	Bk	-	-	Generalized Linear Model (GLM) - binomial: Logistic Regression (Binomial Family)
m9	-	Mk	-	Generalized Linear Model (GLM) - multinomial: Multiclass Classification (Multinomial Family)

3 Итоговая таблица моделей

Таблица 4. Сводная таблица с результатами экспериментов.

Код модели/ Имя модели/файла	№ экс.	Задача - Классиф. бинарная (Bk) - Классиф. множ. (Mk) - Регрессия (Rr)	Метод МО	Набор входных признаков	MSE	RMSE	r2	Max F1	UUID (ИД модели в БЗ)
05_01_Bk_m1_MTarg_F	1	Классиф. бинарная	GBM	Полный	0.010497	0.102457	0.939127	1	ed57d07d-42e3-492a-b202-e2fdc096f9c1
05_02_Bk_m1_MTarg_S	2	Классиф. бинарная	GBM	x87, x74, x76, x72, x70, x73, x68, x67, x81, x80	0.013646	0.116817	0.920868	1	914b816d-6798-4d2a-8e98-02269adb6796
05_03_Bk_m8_MTarg_F	3	Классиф. бинарная	GLM	Полный	0.026021	0.161310	0.847954	0.9912	5f0cbd17-122b-480b-b467-8fbc31bf3741
05_04_Bk_m8_MTarg_S	4	Классиф. бинарная	GLM	x76, x73, x67, x74, x87, x68, x86, x79, x72, x71, x78, x69, x85, x28, x88	0.026837	0.16382	0.843185	0.9869	f3c8cb8c-3218-4f41-903e-af73d5fc7ed8
05_05_Mk_m1_Targ3m_F	5	Классификация множ.	GBM	Полный	0.024655	0.157020	0.958415	-	1cfd41f7-f520-49e3-9e5f-b5c3bd182fa8
05_06_Mk_m1_Targ3m_S	6	Классификация множ.	GBM	x76, x87, x86, x75, x74, x89, x73, x88, x24, x71, x72, x85, x4, x12, x67	0.029854	0.172783	0.949647	-	32b74501-66ed-4f13-bb00-b2a534f37eb8
05_07_Mk_m9_Targ3m_F	7	Классификация множ.	GLM	Полный	0.138441	0.372077	0.771398	-	150a9b8d-7ee9-4015-a3e7-c977c5fe753b
05_08_Mk_m9_Targ3m_S	8	Классификация множ.	GLM	x76, x73, x75, x86, x67, x74, x90, x10, x89, x79, x87, x28, x68, x88, x70	0.145085	0.380900	0.760428	-	9d6449da-876e-4820-9383-ae00cdac38e2
05_09_Rr_m4_Targ3r_F	9	Регрессия (gaussian)	GLM	Полный	0.164396	0.405458	0.728540	-	70056ab0-84f5-4669-947f-eb9e366f644c
05_10_Rr_m4_Targ3r_S	10	Регрессия (gaussian)	GLM	x76, x73, x87, x67, x23, x79, x68, x88, x85, x75, x28	0.166716	0.408309	0.724710	-	0f48a4ae-3615-4d97-8bf6-5c728d20c7ac
05_11_Rr_m7_Targ3r_F	11	Регрессия (gamma)	GLM	Полный	0.219778	0.468803	0.637091	-	f1d64873-a9ea-442d-b022-d6287e41187f
05_12_Rr_m7_Targ3r_S	12	Регрессия (gamma)	GLM	x76, x73, x23, x85, x67, x68, x87, x79, x86, x25	0.219812	0.468841	0.637036	-	d2be50f4-1dee-4a67-814d-c78e607574d7

4 Ход работы

4.1 Загрузка исходных данных в систему H2O

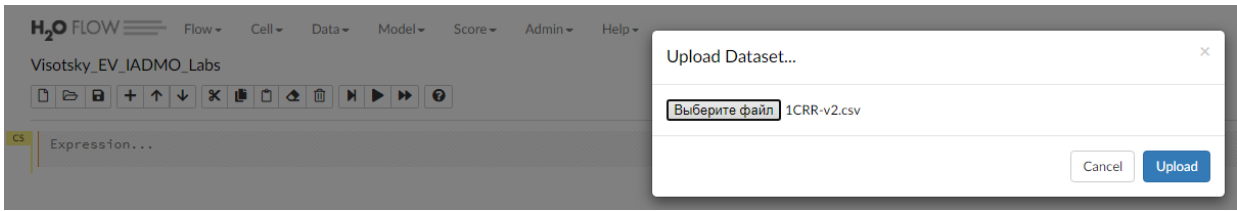


Рис 1. Загрузка файла исходных данных «1CRR-v2.csv» в систему H2O.

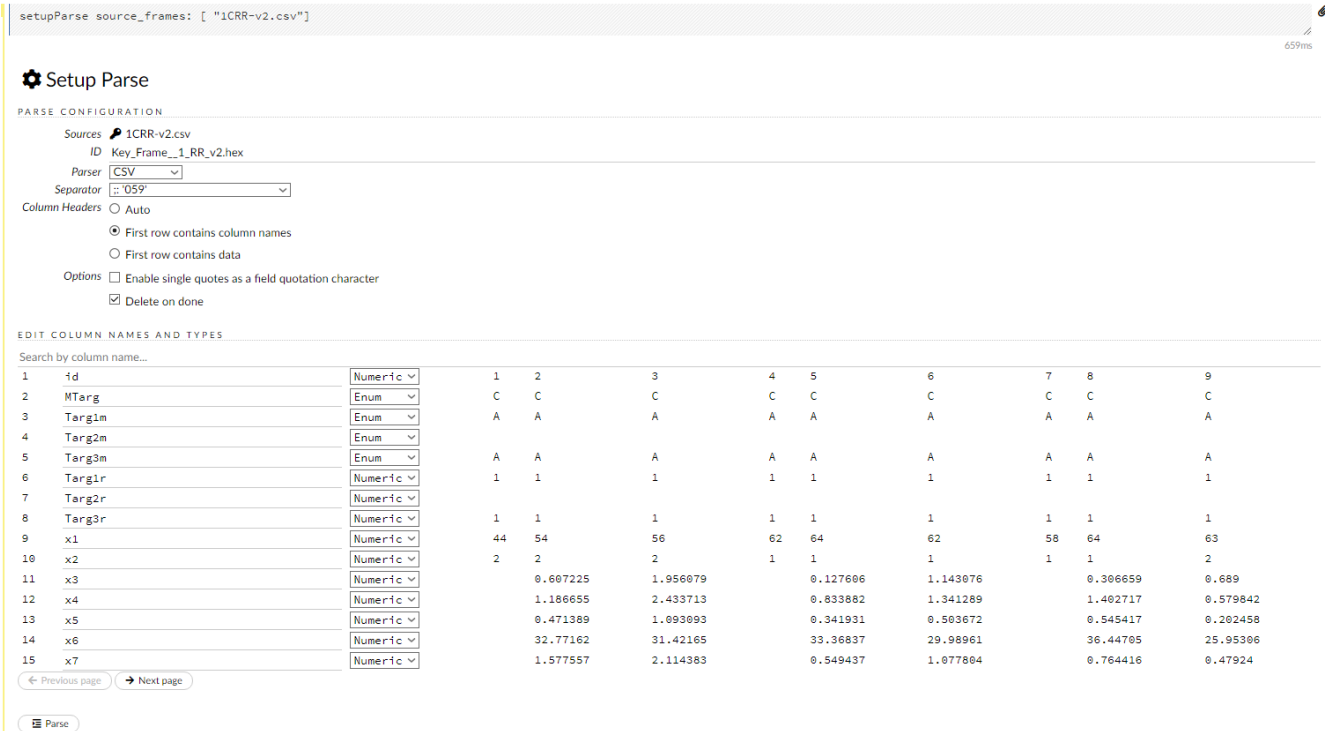


Рис 2. Проверка типов целевых параметров.

4.2 Парсинг загруженных в систему данных

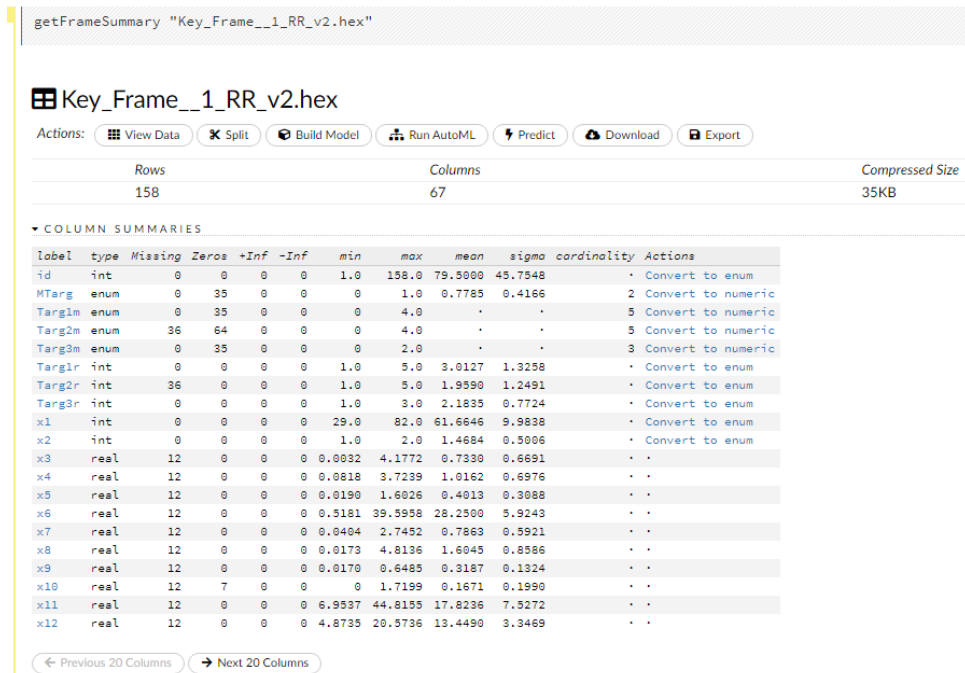


Рис 3. Файл «Key_Frame__1_RR_v2.hex» с распознанными исходными данными.

4.3 Эксперимент № 1. Метод Y1 – Gradient Boosting Machine (GBM) для полной модели «05_01_Bk_MTarg_m1_F»

4.3.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_01_Bk_MTarg_m1_F_0.750", "05_01_Bk_MTarg_m1_F_0.250"], 9
```

Split Frames

Type Key

05_01_Bk_MTarg_m1_F_0.750

05_01_Bk_MTarg_m1_F_0.250

Рис 4. Фреймы данных для модели.

4.3.2 Построение полной модели «05_01_Bk_m1_MTarg_F»

```
buildModel "05_01_Bk_MTarg_m1_F"
```

Build a Model

Select an algorithm: Gradient Boosting Machine

PARAMETERS

model_ida427e15a-1dd7-4828-b409-be3fa87b6a87

training_frame05_01_Bk_MTarg_m1_F_0.750

validation_frameKey_Frame__1_RR_v2.hex

nolds0

response_columnMTarg

ignored_columnsSearch...

Showing page 1 of 7. -60 ignored.

☒id

INT

☐MTarg

ENUM(2)

☒Targ1m

ENUM(5)

☒Targ2m

ENUM(5)25% NA

☒Targ3m

ENUM(3)

☒Targ1r

INT

☒Targ2r

INT25% NA

☒Targ3r

INT

☐x1

INT

☐x2

INT

☒All

☐None

Previous 10

Next 10

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Response variable column.

Names of columns to ignore for training.

Рис 5. Настройка параметров модели.

Job

Run Time00:00:01.967

Remaining Time00:00:00.0

TypeModel

Keya427e15a-1dd7-4828-b409-be3fa87b6a87

DescriptionGBM

StatusDONE

Progress100%

Done.

ActionsView

Рис 6. Отчёт о создании модели.

4.4 Эксперимент № 2. Метод Y1 – Gradient Boosting Machine (GBM) для усечённой модели «05_02_Bk_m1_MTarg_S»

4.4.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_02_Bk_m1_MTarg_S_0.750", "05_02_Bk_m1_MTarg_S_0.250"], 9
```

Split Frames

Type Key

- 05_02_Bk_m1_MTarg_S_0.750
- 05_02_Bk_m1_MTarg_S_0.250

Рис 7. Фреймы данных для модели.

4.4.2 Построение усечённой модели «05_02_Bk_m1_MTarg_S»

```
buildModel "05_02_Bk_m1_MTarg_S"
```

Build a Model

Select an algorithm: Gradient Boosting Machine

PARAMETERS

model_id fba73bb7-d8af-4684-a257-1d7e2dbb70c8

training_frame 05_02_Bk_m1_MTarg_S_0.750

validation_frame Key_Frame__1_RR_v2.hex

nfolds 0

response_column MTarg

ignored_columns Search...

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Response variable column.

Names of columns to ignore for training.

Showing page 1 of 7. 56 ignored.

<input checked="" type="checkbox"/>	id	INT
<input type="checkbox"/>	MTarg	ENUM(2)
<input checked="" type="checkbox"/>	Targ1m	ENUM(5)
<input checked="" type="checkbox"/>	Targ2m	ENUM(5) 25% NA
<input checked="" type="checkbox"/>	Targ3m	ENUM(3)
<input checked="" type="checkbox"/>	Targ1r	INT
<input checked="" type="checkbox"/>	Targ2r	INT 25% NA
<input checked="" type="checkbox"/>	Targ3r	INT
<input checked="" type="checkbox"/>	x1	INT
<input checked="" type="checkbox"/>	x2	INT

☒ All ☐ None Previous 10 Next 10

Рис 8. Настройка параметров модели.

Job

Run Time 00:00:01.455

Remaining Time 00:00:00.0

Type Model

Key fba73bb7-d8af-4684-a257-1d7e2dbb70c8

Description GBM

Status DONE

Progress 100%

Done.

Actions

Рис 9. Отчёт о создании модели.

4.5 Эксперимент № 3. Метод Y1 – Generalized Linear Model (GLM) - binomial: Logistic Regression для полной модели «05_03_Bk_m8_MTarg_F»

4.5.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_03_Bk_m8_MTarg_F_0.750", "05_03_Bk_m8_MTarg_F_0.250"], 9
```

Split Frames

Type Key

05_03_Bk_m8_MTarg_F_0.750

05_03_Bk_m8_MTarg_F_0.250

Рис 10. Фреймы данных для модели.

4.5.2 Построение полной модели «05_03_Bk_m8_MTarg_F»

buildModel "05_03_Bk_m8_MTarg_F"

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id

c84d1e7f-18b2-4bb6-b999-9149f34e5d62

training_frame

05_03_Bk_m8_MTarg_F_0.750

validation_frame

Key_Frame__1_RR_v2.hex

nfolds

0

seed

-1

response_column

MTarg

ignored_columns

Search...

Showing page 1 of 7. -60 ignored.

☒

id

INT

☐

MTarg

ENUM(2)

☒

Targ1m

ENUM(5)

☒

Targ2m

ENUM(5) 25% NA

☒

Targ3m

ENUM(3)

☒

Targ1r

INT

☒

Targ2r

INT 25% NA

☒

Targ3r

INT

☐

x1

INT

☐

x2

INT

☒ All

☐ None

← Previous 10

→ Next 10

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Seed for pseudo random number generator (if applicable)

Response variable column.

Names of columns to ignore for training.

family

binomial

rand_family

Search...

☐ [gaussian]

☒ All

☐ None

solver

AUTO

EXPERT

missing_values_handling

Skip

Рис 11. Настройка параметров модели.

Job

Run Time

00:00:00.235

Remaining Time

00:00:00.0

Type

Model

Key

c84d1e7f-18b2-4bb6-b999-9149f34e5d62

Description

GLM

Status

DONE

Progress

100%

Done.

Actions

View

Рис 12. Отчёт о создании модели.

4.6 Эксперимент № 4. Метод Y1 – Generalized Linear Model (GLM) - binomial: Logistic Regression для усечённой модели «05_04_Bk_m8_MTarg_S»

4.6.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_04_Bk_m8_MTarg_S_0.750", "05_04_Bk_m8_MTarg_S_0.250"], 9
```

Split Frames

Type Key

- 05_04_Bk_m8_MTarg_S_0.750
- 05_04_Bk_m8_MTarg_S_0.250

Рис 13. Фреймы данных для модели.

4.6.2 Построение усечённой модели «05_04_Bk_m8_MTarg_S»

buildModel "05_04_Bk_m8_MTarg_S"

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id72cccf4a-43da-4f63-bc34-2189b9a994c2

training_frame05_04_Bk_m8_MTarg_S_0.750

validation_frameKey_Frame__1_RR_v2.hex

nfolds0

seed-1

response_columnMTarg

ignored_columnsSearch...

Showing page 1 of 7. 51 ignored.

☒

id

INT

☐

MTarg

ENUM(2)

☒

Targ1m

ENUM(5)

☒

Targ2m

ENUM(5) 25% NA

☒

Targ3m

ENUM(3)

☒

Targ1r

INT

☒

Targ2r

INT 25% NA

☒

Targ3r

INT

☒

x1

INT

☒

x2

INT

☒ All

☐ None

Previous 10

Next 10

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Seed for pseudo random number generator (if applicable)

Response variable column.

Names of columns to ignore for training.

familybinomial

rand_familySearch...

☐ [gaussian]

☒ All

☐ None

solverAUTO

EXPERT

missing_values_handlingSkip

Рис 14. Настройка параметров модели.

Job

Run Time00:00:00.239

Remaining Time00:00:00.0

TypeModel

Key72cccf4a-43da-4f63-bc34-2189b9a994c2

DescriptionGLM

StatusDONE

Progress100%

Done.

ActionsView

Рис 15. Отчёт о создании модели.

10

4.7 Эксперимент № 5. Метод Y2 – Gradient Boosting Machine (GBM) для полной модели «05_05_Mk_m1_Targ3m_F»

4.7.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_05_Mk_m1_Targ3m_F_0.750", "05_05_Mk_m1_Targ3m_F_0.250"], 9
```

Split Frames

Type Key

- 05_05_Mk_m1_Targ3m_F_0.750
- 05_05_Mk_m1_Targ3m_F_0.250

Рис 16. Фреймы данных для модели.

4.7.2 Построение полной модели «05_05_Mk_m1_Targ3m_F»

```
buildModel "05_05_Mk_m1_Targ3m_F"
```

Build a Model

Select an algorithm: Gradient Boosting Machine

PARAMETERS

model_id	3230af36-4864-437c-a41d-283191959347	Destination id for this model; auto-generated if not specified.
training_frame	05_05_Mk_m1_Targ3m_F_0.750	Id of the training data frame.
validation_frame	Key_Frame__1_RR_v2.hex	Id of the validation data frame.
nfolds	0	Number of folds for K-fold cross-validation (0 to disable or >= 2).
response_column	Targ3m	Response variable column.
ignored_columns	Search...	Names of columns to ignore for training.

Showing page 1 of 7. -60 ignored.

<input checked="" type="checkbox"/> id	INT
<input checked="" type="checkbox"/> MTarg	ENUM(2)
<input checked="" type="checkbox"/> Targ1m	ENUM(5)
<input checked="" type="checkbox"/> Targ2m	ENUM(5) 25% NA
<input type="checkbox"/> Targ3m	ENUM(3)
<input checked="" type="checkbox"/> Targ1r	INT
<input checked="" type="checkbox"/> Targ2r	INT 25% NA
<input checked="" type="checkbox"/> Targ3r	INT
<input type="checkbox"/> x1	INT
<input type="checkbox"/> x2	INT

☒ All ☐ None ← Previous 10 → Next 10

Рис 17. Настройка параметров модели.

Job

Run Time 00:00:03.16

Remaining Time 00:00:00.0

Type Model

Key Q 3230af36-4864-437c-a41d-283191959347

Description GBM

Status DONE

Progress 100%

Done.

Actions View

Рис 18. Отчёт о создании модели.

4.8 Эксперимент № 6. Метод Y2 – Gradient Boosting Machine (GBM) для усечённой модели «05_06_Mk_m1_Targ3m_S»

4.8.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_06_Mk_m1_Targ3m_S_0.750", "05_06_Mk_m1_Targ3m_S_0.250"], 9
```

Split Frames

Type Key

- 05_06_Mk_m1_Targ3m_S_0.750
- 05_06_Mk_m1_Targ3m_S_0.250

Рис 19. Фреймы данных для модели.

4.8.2 Построение усечённой модели «05_06_Mk_m1_Targ3m_S»

```
buildModel "05_06_Mk_m1_Targ3m_S"
```

Build a Model

Select an algorithm: Gradient Boosting Machine

PARAMETERS

model_id	2f140483-71d8-460b-b36a-3ff1d48c8363	Destination id for this model; auto-generated if not specified.
training_frame	05_06_Mk_m1_Targ3m_S_0.750	Id of the training data frame.
validation_frame	Key_Frame__1_RR_v2.hex	Id of the validation data frame.
nfolds	0	Number of folds for K-fold cross-validation (0 to disable or >= 2).
response_column	Targ3m	Response variable column.
ignored_columns	Search...	Names of columns to ignore for training.

Showing page 1 of 7. 51 ignored.

<input checked="" type="checkbox"/>	id	INT
<input checked="" type="checkbox"/>	MTarg	ENUM(2)
<input checked="" type="checkbox"/>	Targ1m	ENUM(5)
<input checked="" type="checkbox"/>	Targ2m	ENUM(5) 25% NA
<input type="checkbox"/>	Targ3m	ENUM(3)
<input checked="" type="checkbox"/>	Targ1r	INT
<input checked="" type="checkbox"/>	Targ2r	INT 25% NA
<input checked="" type="checkbox"/>	Targ3r	INT
<input checked="" type="checkbox"/>	x1	INT
<input checked="" type="checkbox"/>	x2	INT

☒ All ☐ None ← Previous 10 → Next 10

Рис 20. Настройка параметров модели.

Job

Run Time 00:00:00.984

Remaining Time 00:00:00.0

Type Model

Key 2f140483-71d8-460b-b36a-3ff1d48c8363

Description GBM

Status DONE

Progress 100%

Done.

Actions View

Рис 21. Отчёт о создании модели.

4.9 Эксперимент № 7. Метод Y2 – Generalized Linear Model (GLM) - multinomial: Logistic Regression для полной модели «05_07_Mk_m9_Targ3m_F»

4.9.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_07_Mk_m9_Targ3m_F_0.750", "05_07_Mk_m9_Targ3m_F_0.250"], 9
```

Split Frames

Type Key

05_07_Mk_m9_Targ3m_F_0.750

05_07_Mk_m9_Targ3m_F_0.250

Рис 22. Фреймы данных для модели.

4.9.2 Построение полной модели «05_07_Mk_m9_Targ3m_F»

```
buildModel "05_07_Mk_m9_Targ3m_F"
```

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id

bb373d1e-e658-410d-9836-60e434c6764c

training_frame

05_07_Mk_m9_Targ3m_F_0.750

validation_frame

Key_Frame__1_RR_v2.hex

nfolds

0

seed

-1

response_column

Targ3m

ignored_columns

Search...

Showing page 1 of 7. -60 ignored.

☒

id

INT

☒

MTarg

ENUM(2)

☒

Targ1m

ENUM(5)

☒

Targ2m

ENUM(5) 25% NA

☐

Targ3m

ENUM(3)

☒

Targ1r

INT

☒

Targ2r

INT 25% NA

☒

Targ3r

INT

☐

x1

INT

☐

x2

INT

☒ All

☐ None

Previous 10

Next 10

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Seed for pseudo random number generator (if applicable)

Response variable column.

Names of columns to ignore for training.

family

multinomial

rand_family

Search...

☐ [gaussian]

☒ All

☐ None

solver

AUTO

EXPERT

missing_values_handling

Skip

Рис 23. Настройка параметров модели.

Job

Run Time

00:00:00.937

Remaining Time

00:00:00.0

Type

Model

Key

bb373d1e-e658-410d-9836-60e434c6764c

Description

GLM

Status

DONE

Progress

100%

Done.

Actions

View

Рис 24. Отчёт о создании модели.

13

4.10 Эксперимент № 8. Метод Y2 – Generalized Linear Model (GLM) - multinomial: Logistic Regression для усечённой модели «05_08_Mk_m9_Targ3m_S»

4.10.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_08_Mk_m9_Targ3m_S_0.750", "05_08_Mk_m9_Targ3m_S_0.250"], 9
```

Split Frames

Type	Key
	05_08_Mk_m9_Targ3m_S_0.750
	05_08_Mk_m9_Targ3m_S_0.250

Рис 25. Фреймы данных для модели.

4.10.2 Построение усечённой модели «05_08_Mk_m9_Targ3m_S»

```
buildModel "05_08_Mk_m9_Targ3m_S"
```

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id	e3f51438-11bb-4992-a7b2-847d75b2f410	Destination id for this model; auto-generated if not specified.
training_frame	05_08_Mk_m9_Targ3m_S_0.750	Id of the training data frame.
validation_frame	Key_Frame__1_RR_v2.hex	Id of the validation data frame.
nfolds	0	Number of folds for K-fold cross-validation (0 to disable or >= 2).
seed	-1	Seed for pseudo random number generator (if applicable)
response_column	Targ3m	Response variable column.
ignored_columns	Search...	Names of columns to ignore for training.
Showing page 7 of 7. 51 ignored.		
<div><input checked="" type="checkbox"/> x84 REAL</div> <div><input checked="" type="checkbox"/> x85 REAL</div> <div><input type="checkbox"/> x86 REAL</div> <div><input type="checkbox"/> x87 REAL</div> <div><input type="checkbox"/> x88 REAL</div> <div><input type="checkbox"/> x89 REAL</div> <div><input type="checkbox"/> x90 REAL</div>		<div>family multinomial</div> <div>rand_family Search...</div> <div><input type="checkbox"/> [gaussian]</div> <div><input checked="" type="checkbox"/> All <input type="checkbox"/> None</div> <div>solver AUTO</div>
<div><input checked="" type="checkbox"/> All <input type="checkbox"/> None</div> <div>← Previous 10 → Next 10</div>		EXPERT
		missing_values_handling Skip

Рис 26. Настройка параметров модели.

Job

Run Time	00:00:00.315
Remaining Time	00:00:00.0
Type	Model
Key	Q e3f51438-11bb-4992-a7b2-847d75b2f410
Description	GLM
Status	DONE
Progress	100% <div></div>
Done.	
Actions	<div>Q View</div>

Рис 27. Отчёт о создании модели.

4.11 Эксперимент № 9. Метод Y3 – Generalized Linear Model (GLM) - gaussian: Gaussian regression для полной модели «05_09_Rr_m4_Targ3r_F»

4.11.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_09_Rr_m4_Targ3r_F_0.750","05_09_Rr_m4_Targ3r_F_0.250"], 9
```

Split Frames

Type Key

05_09_Rr_m4_Targ3r_F_0.750

05_09_Rr_m4_Targ3r_F_0.250

Рис 28. Фреймы данных для модели.

4.11.2 Построение полной модели «05_09_Rr_m4_Targ3r_F»

```
buildModel "05_09_Rr_m4_Targ3r_F"
```

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id

e4c558a7-e30c-4aec-a858-5fa0e29a9429

training_frame

05_09_Rr_m4_Targ3r_F_0.750

validation_frame

Key_Frame__1_RR_v2.hex

nfolds

0

seed

-1

response_column

Targ3r

ignored_columns

Search...

Showing page 1 of 7. -60 ignored.

☒

id

INT

☒

MTarg

ENUM(2)

☒

Targ1m

ENUM(5)

☒

Targ2m

ENUM(5) 25% NA

☒

Targ3m

ENUM(3)

☒

Targ1r

INT

☒

Targ2r

INT 25% NA

☐

Targ3r

INT

☐

x1

INT

☐

x2

INT

☒ All

☐ None

← Previous 10

→ Next 10

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Seed for pseudo random number generator (if applicable)

Response variable column.

Names of columns to ignore for training.

family

gaussian

EXPERT

missing_values_handling

Skip

Рис 29. Настройка параметров модели.

Job

Run Time

00:00:00.275

Remaining Time

00:00:00.0

Type

Model

Key

e4c558a7-e30c-4aec-a858-5fa0e29a9429

Description

GLM

Status

DONE

Progress

100%

Done.

Actions

View

Рис 30. Отчёт о создании модели.

4.12 Эксперимент № 10. Метод Y3 – Generalized Linear Model (GLM) - gaussian: Gaussian regression для усечённой модели «05_10_Rr_m4_Targ3r_S»

4.12.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_10_Rr_m4_Targ3r_S_0.750", "05_10_Rr_m4_Targ3r_S_0.250"], 9
```



Рис 31. Фреймы данных для модели.

4.12.2 Построение усечённой модели «05_10_Rr_m4_Targ3r_S»

```
buildModel "05_10_Rr_m4_Targ3r_S"
```

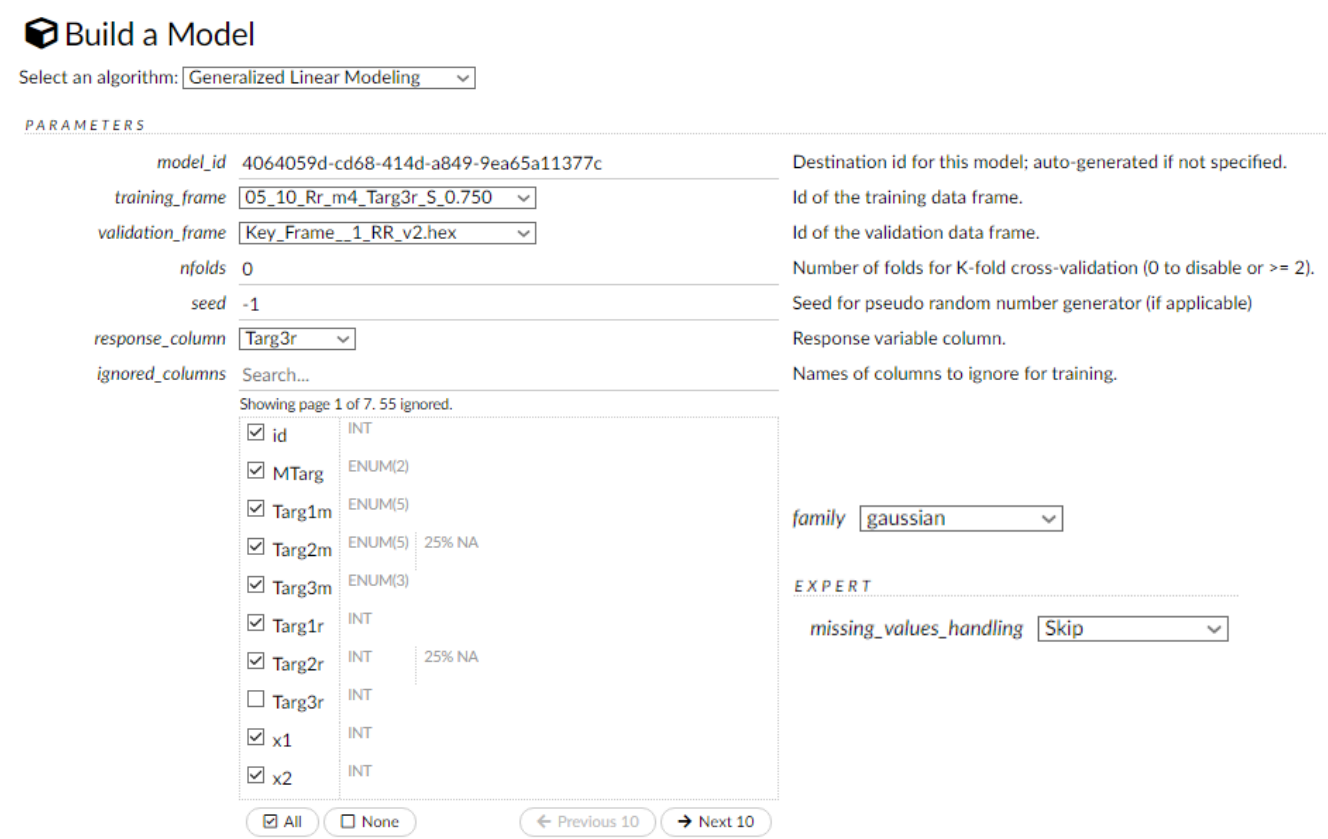


Рис 32. Настройка параметров модели.

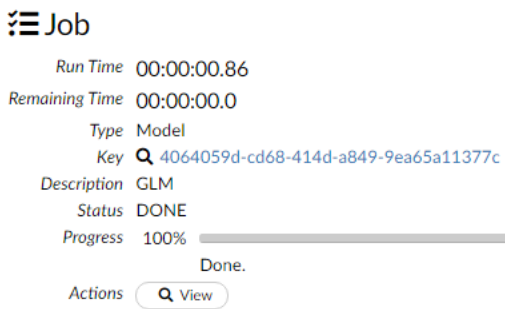


Рис 33. Отчёт о создании модели.

4.13 Эксперимент № 11. Метод Y3 – Generalized Linear Model (GLM) - gamma: Gamma Models для полной модели «05_11_Rr_m7_Targ3r_F»

4.13.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_11_Rr_m7_Targ3r_F_0.750", "05_11_Rr_m7_Targ3r_F_0.250"], 9
```

Split Frames

Type Key

05_11_Rr_m7_Targ3r_F_0.750

05_11_Rr_m7_Targ3r_F_0.250

Рис 34. Фреймы данных для модели.

4.13.2 Построение полной модели «05_11_Rr_m7_Targ3r_F»

```
buildModel "05_11_Rr_m7_Targ3r_F"
```

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id

e36c6dd6-20de-4640-b993-263841216a34

training_frame

05_11_Rr_m7_Targ3r_F_0.750

validation_frame

Key_Frame__1_RR_v2.hex

nfolds

0

seed

-1

response_column

Targ3r

ignored_columns

Search...

Destination id for this model; auto-generated if not specified.

Id of the training data frame.

Id of the validation data frame.

Number of folds for K-fold cross-validation (0 to disable or >= 2).

Seed for pseudo random number generator (if applicable)

Response variable column.

Names of columns to ignore for training.

Showing page 1 of 7. -60 ignored.

☒

id

INT

☒

MTarg

ENUM(2)

☒

Targ1m

ENUM(5)

☒

Targ2m

ENUM(5) 25% NA

☒

Targ3m

ENUM(3)

☒

Targ1r

INT

☒

Targ2r

INT 25% NA

☐

Targ3r

INT

☐

x1

INT

☐

x2

INT

☒ All

☐ None

Previous 10

Next 10

Рис 35. Настройка параметров модели.

Job

Run Time

00:00:00.338

Remaining Time

00:00:00.0

Type

Model

Key

e36c6dd6-20de-4640-b993-263841216a34

Description

GLM

Status

DONE

Progress

100%

Done.

Actions

View

Рис 36. Отчёт о создании модели.

4.14 Эксперимент № 12. Метод Y3 – Generalized Linear Model (GLM) - gamma: Gamma Models для усечённой модели «05_10_Rr_m4_Targ3r_S»

4.14.1 Разбиение данных для обучения и проверки модели

```
splitFrame "Key_Frame__1_RR_v2.hex", [0.75], ["05_12_Rr_m7_Targ3r_S_0.750", "05_12_Rr_m7_Targ3r_S_0.250"], 9
```

Split Frames

Type Key

- 05_12_Rr_m7_Targ3r_S_0.750
- 05_12_Rr_m7_Targ3r_S_0.250

Рис 37. Фреймы данных для модели.

4.14.2 Построение усечённой модели «05_10_Rr_m4_Targ3r_S»

```
buildModel "05_12_Rr_m7_Targ3r_S"
```

Build a Model

Select an algorithm: Generalized Linear Modeling

PARAMETERS

model_id	3ac922a4-4679-4de3-a12e-a7fc22031b0c	Destination id for this model; auto-generated if not specified.
training_frame	05_12_Rr_m7_Targ3r_S_0.750	Id of the training data frame.
validation_frame	Key_Frame__1_RR_v2.hex	Id of the validation data frame.
nfolds	0	Number of folds for K-fold cross-validation (0 to disable or >= 2).
seed	-1	Seed for pseudo random number generator (if applicable)
response_column	Targ3r	Response variable column.
ignored_columns	Search...	Names of columns to ignore for training.

Showing page 1 of 7. 56 ignored.

<input checked="" type="checkbox"/>	id	INT
<input checked="" type="checkbox"/>	MTarg	ENUM(2)
<input checked="" type="checkbox"/>	Targ1m	ENUM(5)
<input checked="" type="checkbox"/>	Targ2m	ENUM(5) 25% NA
<input checked="" type="checkbox"/>	Targ3m	ENUM(3)
<input checked="" type="checkbox"/>	Targ1r	INT
<input checked="" type="checkbox"/>	Targ2r	INT 25% NA
<input type="checkbox"/>	Targ3r	INT
<input checked="" type="checkbox"/>	x1	INT
<input checked="" type="checkbox"/>	x2	INT

☒ All ☐ None

← Previous 10 Next 10 →

family gamma

EXPERT

missing_values_handling Skip

Рис 38. Настройка параметров модели.

Job

Run Time 00:00:00.141

Remaining Time 00:00:00.0

Type Model

Key 3ac922a4-4679-4de3-a12e-a7fc22031b0c

Description GLM

Status DONE

Progress 100%

Done.

Actions View

Рис 39. Отчёт о создании модели.

4.15 Анализ информации о созданных моделях

4.15.1 Важность признаков в графическом виде

Таблица 5. Важность признаков моделей.

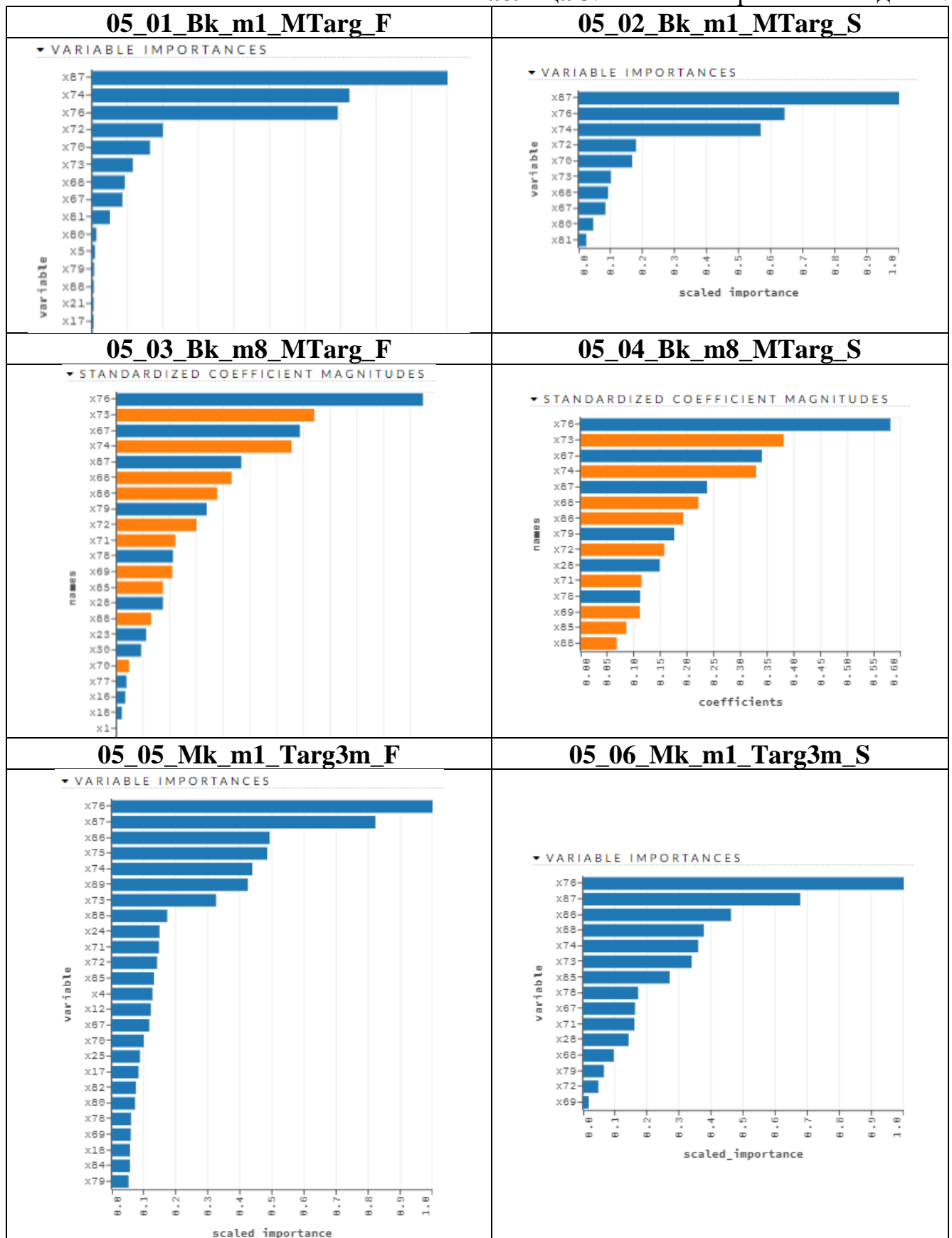
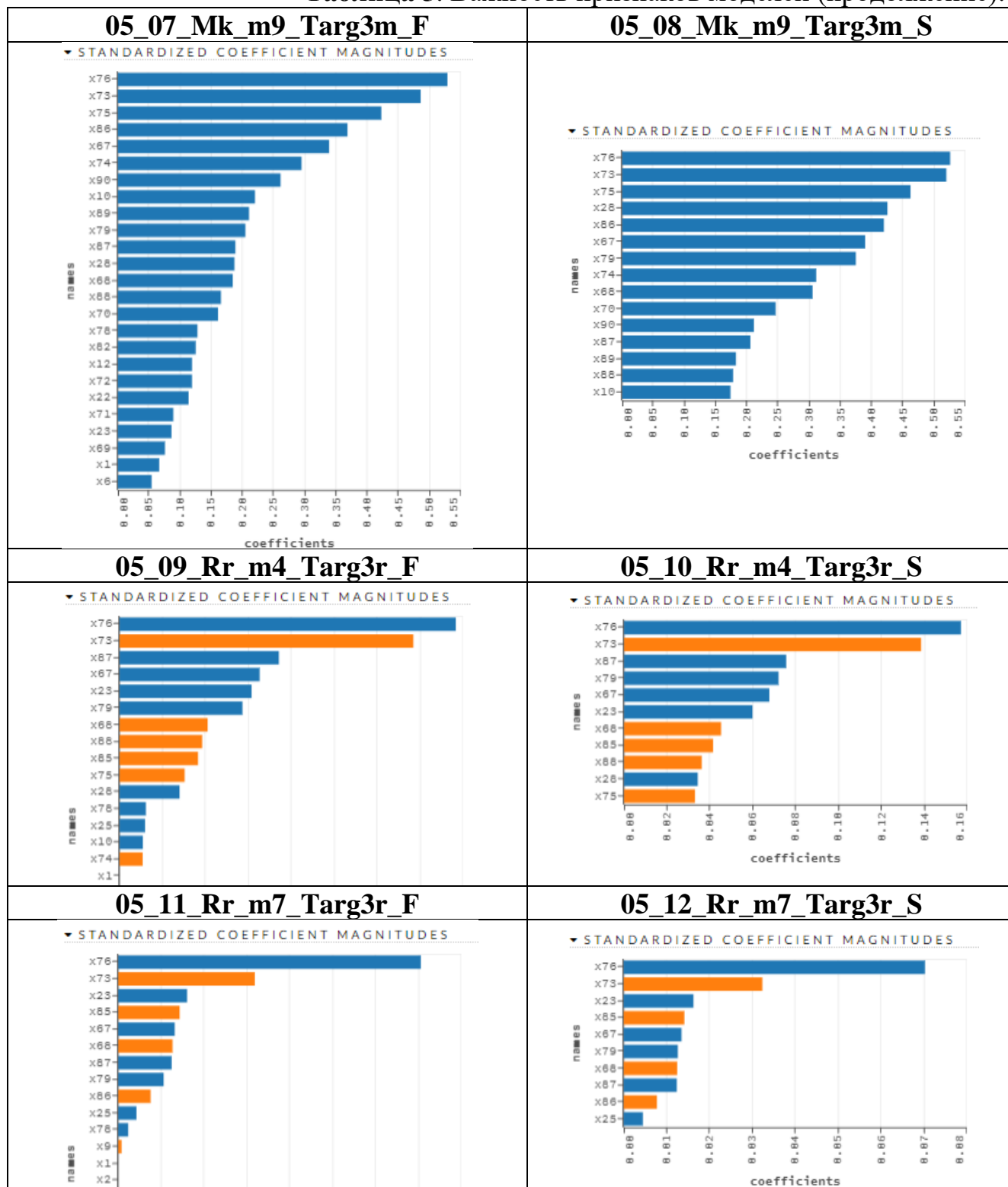


Таблица 5. Важность признаков моделей (продолжение).



4.15.2 Важность признаков в числовом виде

Таблица 6. Сравнение признаков моделей в числовом виде.

05_01_Bk_m1_MTarg_F					05_02_Bk_m1_MTarg_S				
variable	relative_importance	scaled_importance	percenta	^	variable	relative_importance	scaled_importance	percentage	
x87	36.7900	1.0	0.31		x87	40.2597	1.0	0.3448	
x74	26.6001	0.7230	0.22		x76	25.8390	0.6418	0.2213	
x76	25.4015	0.6904	0.21		x74	22.8442	0.5674	0.1957	
x72	7.3014	0.1985	0.06		x72	7.1993	0.1788	0.0617	
x70	5.9649	0.1621	0.05		x70	6.7108	0.1667	0.0575	
x73	4.2000	0.1142	0.03		x73	4.0490	0.1006	0.0347	
x68	3.3680	0.0915	0.02		x68	3.6944	0.0918	0.0316	
x67	3.1145	0.0847	0.02		x67	3.3670	0.0836	0.0288	
x81	1.8278	0.0497	0.01		x80	1.8296	0.0454	0.0157	
x80	0.4268	0.0116	0.00		x81	0.9592	0.0238	0.0082	
x5	0.2770	0.0075	0.00						
05_03_Bk_m8_MTarg_F					05_04_Bk_m8_MTarg_S				
variable	relative_importance	scaled_importance	percenta	^	variable	relative_importance	scaled_importance	percentage	
x76	0.5725	1.0	0.17		x76	0.5797	1.0	0.1787	
x73	0.3695	0.6454	0.11		x73	0.3800	0.6556	0.1171	
x67	0.3425	0.5982	0.10		x67	0.3393	0.5853	0.1046	
x74	0.3270	0.5712	0.09		x74	0.3284	0.5665	0.1012	
x87	0.2331	0.4071	0.07		x87	0.2366	0.4081	0.0729	
x68	0.2149	0.3754	0.06		x68	0.2205	0.3804	0.0680	
x86	0.1879	0.3282	0.05		x86	0.1920	0.3311	0.0592	
x79	0.1683	0.2939	0.05		x79	0.1750	0.3019	0.0539	
x72	0.1492	0.2605	0.04		x72	0.1566	0.2702	0.0483	
x71	0.1099	0.1919	0.03		x28	0.1478	0.2550	0.0456	
x78	0.1052	0.1838	0.03		x71	0.1138	0.1964	0.0351	
x69	0.1040	0.1816	0.03		x78	0.1113	0.1920	0.0343	
x85	0.0865	0.1511	0.02		x69	0.1107	0.1909	0.0341	
x28	0.0864	0.1510	0.02		x85	0.0856	0.1477	0.0264	
x88	0.0643	0.1123	0.01		x88	0.0674	0.1162	0.0208	
05_05_Mk_m1_Targ3m_F					05_06_Mk_m1_Targ3m_S				
variable	relative_importance	scaled_importance	percenta	^	variable	relative_importance	scaled_importance	percentage	
x76	55.4540	1.0	0.15		x76	81.8404	1.0	0.2305	
x87	45.5119	0.8207	0.12		x87	55.3614	0.6765	0.1560	
x86	27.2492	0.4914	0.07		x86	37.6855	0.4605	0.1062	
x75	26.8236	0.4837	0.07		x88	30.7696	0.3760	0.0867	
x74	24.2462	0.4372	0.06		x74	29.3175	0.3582	0.0826	
x89	23.4795	0.4234	0.06		x73	27.6745	0.3382	0.0780	
x73	17.9970	0.3245	0.05		x85	22.0601	0.2696	0.0621	
x88	9.5550	0.1723	0.02		x78	14.0273	0.1714	0.0395	
x24	8.2210	0.1482	0.02		x67	13.2183	0.1615	0.0372	
x71	8.1288	0.1466	0.02		x71	13.0628	0.1596	0.0368	
x72	7.8006	0.1407	0.02		x28	11.5766	0.1415	0.0326	
x85	7.2709	0.1311	0.02		x68	7.8014	0.0953	0.0220	
x4	7.0078	0.1264	0.01		x79	5.2926	0.0647	0.0149	
x12	6.7335	0.1214	0.01		x72	3.8764	0.0474	0.0109	
					x69	1.4166	0.0173	0.0040	
05_07_Mk_m9_Targ3m_F					05_08_Mk_m9_Targ3m_S				
variable	relative_importance	scaled_importance	percenta	^	variable	relative_importance	scaled_importance	percentage	
x76	0.5277	1.0	0.09		x76	0.5249	1.0	0.1067	
x73	0.4848	0.9187	0.09		x73	0.5188	0.9884	0.1054	
x75	0.4219	0.7994	0.07		x75	0.4614	0.8791	0.0938	
x86	0.3677	0.6967	0.06		x28	0.4245	0.8088	0.0863	
x67	0.3383	0.6411	0.06		x86	0.4189	0.7980	0.0851	
x74	0.2941	0.5572	0.05		x67	0.3889	0.7409	0.0790	
x90	0.2605	0.4936	0.04		x79	0.3740	0.7125	0.0760	
x10	0.2195	0.4159	0.04		x74	0.3105	0.5915	0.0631	
x89	0.2100	0.3979	0.03		x68	0.3047	0.5804	0.0619	
x79	0.2042	0.3869	0.03		x70	0.2456	0.4678	0.0499	
x87	0.1881	0.3565	0.03		x90	0.2108	0.4016	0.0428	
x28	0.1866	0.3536	0.03		x87	0.2052	0.3909	0.0417	
x68	0.1838	0.3484	0.03		x89	0.1821	0.3470	0.0370	
x88	0.1650	0.3126	0.03		x88	0.1774	0.3381	0.0361	
x70	0.1602	0.3036	0.02		x10	0.1732	0.3300	0.0352	

Таблица 6. Сравнение признаков моделей в числовом виде (продолжение).

05_09_Rr_m4_Targ3r_F				05_10_Rr_m4_Targ3r_S			
variable relative_importance scaled_importance percenta ^				variable relative_importance scaled_importance percentage			
x76	0.1564	1.0	0.20	x76	0.1569	1.0	0.2063
x73	0.1367	0.8738	0.17	x73	0.1383	0.8819	0.1820
x87	0.0741	0.4739	0.09	x87	0.0755	0.4812	0.0993
x67	0.0652	0.4166	0.08	x79	0.0719	0.4586	0.0946
x23	0.0614	0.3927	0.07	x67	0.0677	0.4313	0.0890
x79	0.0572	0.3655	0.07	x23	0.0598	0.3812	0.0787
x68	0.0410	0.2621	0.05	x68	0.0452	0.2881	0.0594
x88	0.0384	0.2458	0.04	x85	0.0415	0.2644	0.0545
x85	0.0365	0.2335	0.04	x88	0.0362	0.2306	0.0476
x75	0.0303	0.1937	0.03	x28	0.0343	0.2187	0.0451
x28	0.0279	0.1785	0.03	x75	0.0330	0.2105	0.0434
x78	0.0123	0.0783	0.01				
x25	0.0119	0.0758	0.01				
x10	0.0108	0.0692	0.01				
x74	0.0107	0.0686	0.01				

05_11_Rr_m7_Targ3r_F				05_12_Rr_m7_Targ3r_S			
variable relative_importance scaled_importance percentage ^				variable relative_importance scaled_importance percentage			
x87	0.0124	0.1765	0.06	x76	0.0702	1.0	0.3582
x79	0.0106	0.1500	0.05	x73	0.0323	0.4602	0.1648
x86	0.0075	0.1069	0.03	x23	0.0162	0.2316	0.0829
x25	0.0042	0.0600	0.02	x85	0.0141	0.2013	0.0721
x78	0.0023	0.0328	0.01	x67	0.0135	0.1919	0.0687
x9	0.0008	0.0108	0.00	x79	0.0126	0.1796	0.0643
x1	0	0		x68	0.0125	0.1775	0.0636
				x87	0.0124	0.1761	0.0631
				x86	0.0077	0.1102	0.0395
				x25	0.0045	0.0636	0.0228

4.15.3 Матрицы ошибок

Таблица 7. Оценка ошибок моделирования.

Оценка на обучающем наборе

Оценка на полном наборе

05_01_Bk_m1_MTarg_F

TRAINING METRICS - CONFUSION MATRIX

	C	F	Error	Rate	Precision
C	30	0	0	0 / 30	1.0
F	0	95	0	0 / 95	1.0
Total	30	95	0	0 / 125	
Recall	1.0	1.0			

VALIDATION METRICS - CONFUSION MATRIX

	C	F	Error	Rate	Precision
C	35	0	0	0 / 35	1.0
F	0	123	0	0 / 123	1.0
Total	35	123	0	0 / 158	
Recall	1.0	1.0			

05_02_Bk_m1_MTarg_S

TRAINING METRICS - CONFUSION MATRIX

	C	F	Error	Rate	Precision
C	30	0	0	0 / 30	1.0
F	0	95	0	0 / 95	1.0
Total	30	95	0	0 / 125	
Recall	1.0	1.0			

VALIDATION METRICS - CONFUSION MATRIX

	C	F	Error	Rate	Precision
C	35	0	0	0 / 35	1.0
F	0	123	0	0 / 123	1.0
Total	35	123	0	0 / 158	
Recall	1.0	1.0			

Таблица 7. Оценка ошибок моделирования (продолжение).

Оценка на обучающем наборе		Оценка на полном наборе																																																																																					
05_03_Bk_m8_MTarg_F																																																																																							
HET		<div>▼ PREDICTION - CONFUSION MATRIX ROW L</div> <table><tr><td></td><td>C</td><td>F</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>C</td><td>32</td><td>0</td><td>0</td><td>0 / 32</td><td>0.94</td></tr><tr><td>F</td><td>2</td><td>112</td><td>0.0175</td><td>2 / 114</td><td>1.0</td></tr><tr><td>Total</td><td>34</td><td>112</td><td>0.0137</td><td>2 / 146</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>0.98</td><td></td><td></td><td></td></tr></table>			C	F	Error	Rate	Precision	C	32	0	0	0 / 32	0.94	F	2	112	0.0175	2 / 114	1.0	Total	34	112	0.0137	2 / 146		Recall	1.0	0.98																																																									
	C	F	Error	Rate	Precision																																																																																		
C	32	0	0	0 / 32	0.94																																																																																		
F	2	112	0.0175	2 / 114	1.0																																																																																		
Total	34	112	0.0137	2 / 146																																																																																			
Recall	1.0	0.98																																																																																					
05_04_Bk_m8_MTarg_S																																																																																							
HET		<div>▼ PREDICTION - CONFUSION MATRIX ROW LAI</div> <table><tr><td></td><td>C</td><td>F</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>C</td><td>30</td><td>2</td><td>0.0625</td><td>2 / 32</td><td>0.97</td></tr><tr><td>F</td><td>1</td><td>113</td><td>0.0088</td><td>1 / 114</td><td>0.98</td></tr><tr><td>Total</td><td>31</td><td>115</td><td>0.0205</td><td>3 / 146</td><td></td></tr><tr><td>Recall</td><td>0.94</td><td>0.99</td><td></td><td></td><td></td></tr></table>			C	F	Error	Rate	Precision	C	30	2	0.0625	2 / 32	0.97	F	1	113	0.0088	1 / 114	0.98	Total	31	115	0.0205	3 / 146		Recall	0.94	0.99																																																									
	C	F	Error	Rate	Precision																																																																																		
C	30	2	0.0625	2 / 32	0.97																																																																																		
F	1	113	0.0088	1 / 114	0.98																																																																																		
Total	31	115	0.0205	3 / 146																																																																																			
Recall	0.94	0.99																																																																																					
05_05_Mk_m1_Targ3m_F																																																																																							
<div>▼ TRAINING METRICS - CONFUSION MATRIX ROW</div> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>A</td><td>30</td><td>0</td><td>0</td><td>0</td><td>0 / 30</td><td>1.0</td></tr><tr><td>B</td><td>0</td><td>46</td><td>0</td><td>0</td><td>0 / 46</td><td>1.0</td></tr><tr><td>C</td><td>0</td><td>0</td><td>49</td><td>0</td><td>0 / 49</td><td>1.0</td></tr><tr><td>Total</td><td>30</td><td>46</td><td>49</td><td>0</td><td>0 / 125</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>1.0</td><td>1.0</td><td></td><td></td><td></td></tr></table>			A	B	C	Error	Rate	Precision	A	30	0	0	0	0 / 30	1.0	B	0	46	0	0	0 / 46	1.0	C	0	0	49	0	0 / 49	1.0	Total	30	46	49	0	0 / 125		Recall	1.0	1.0	1.0				<div>▼ VALIDATION METRICS - CONFUSION MATRIX ROW</div> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>A</td><td>35</td><td>0</td><td>0</td><td>0</td><td>0 / 35</td><td>0.95</td></tr><tr><td>B</td><td>1</td><td>57</td><td>1</td><td>0.0339</td><td>2 / 59</td><td>0.97</td></tr><tr><td>C</td><td>1</td><td>2</td><td>61</td><td>0.0469</td><td>3 / 64</td><td>0.98</td></tr><tr><td>Total</td><td>37</td><td>59</td><td>62</td><td>0.0316</td><td>5 / 158</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>0.97</td><td>0.95</td><td></td><td></td><td></td></tr></table>			A	B	C	Error	Rate	Precision	A	35	0	0	0	0 / 35	0.95	B	1	57	1	0.0339	2 / 59	0.97	C	1	2	61	0.0469	3 / 64	0.98	Total	37	59	62	0.0316	5 / 158		Recall	1.0	0.97	0.95			
	A	B	C	Error	Rate	Precision																																																																																	
A	30	0	0	0	0 / 30	1.0																																																																																	
B	0	46	0	0	0 / 46	1.0																																																																																	
C	0	0	49	0	0 / 49	1.0																																																																																	
Total	30	46	49	0	0 / 125																																																																																		
Recall	1.0	1.0	1.0																																																																																				
	A	B	C	Error	Rate	Precision																																																																																	
A	35	0	0	0	0 / 35	0.95																																																																																	
B	1	57	1	0.0339	2 / 59	0.97																																																																																	
C	1	2	61	0.0469	3 / 64	0.98																																																																																	
Total	37	59	62	0.0316	5 / 158																																																																																		
Recall	1.0	0.97	0.95																																																																																				
05_06_Mk_m1_Targ3m_S																																																																																							
<div>▼ TRAINING METRICS - CONFUSION MATRIX ROW</div> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>A</td><td>30</td><td>0</td><td>0</td><td>0</td><td>0 / 30</td><td>1.0</td></tr><tr><td>B</td><td>0</td><td>46</td><td>0</td><td>0</td><td>0 / 46</td><td>1.0</td></tr><tr><td>C</td><td>0</td><td>0</td><td>49</td><td>0</td><td>0 / 49</td><td>1.0</td></tr><tr><td>Total</td><td>30</td><td>46</td><td>49</td><td>0</td><td>0 / 125</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>1.0</td><td>1.0</td><td></td><td></td><td></td></tr></table>			A	B	C	Error	Rate	Precision	A	30	0	0	0	0 / 30	1.0	B	0	46	0	0	0 / 46	1.0	C	0	0	49	0	0 / 49	1.0	Total	30	46	49	0	0 / 125		Recall	1.0	1.0	1.0				<div>▼ VALIDATION METRICS - CONFUSION MATRIX ROW</div> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>A</td><td>35</td><td>0</td><td>0</td><td>0</td><td>0 / 35</td><td>0.95</td></tr><tr><td>B</td><td>1</td><td>56</td><td>2</td><td>0.0508</td><td>3 / 59</td><td>0.97</td></tr><tr><td>C</td><td>1</td><td>2</td><td>61</td><td>0.0469</td><td>3 / 64</td><td>0.97</td></tr><tr><td>Total</td><td>37</td><td>58</td><td>63</td><td>0.0380</td><td>6 / 158</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>0.95</td><td>0.95</td><td></td><td></td><td></td></tr></table>			A	B	C	Error	Rate	Precision	A	35	0	0	0	0 / 35	0.95	B	1	56	2	0.0508	3 / 59	0.97	C	1	2	61	0.0469	3 / 64	0.97	Total	37	58	63	0.0380	6 / 158		Recall	1.0	0.95	0.95			
	A	B	C	Error	Rate	Precision																																																																																	
A	30	0	0	0	0 / 30	1.0																																																																																	
B	0	46	0	0	0 / 46	1.0																																																																																	
C	0	0	49	0	0 / 49	1.0																																																																																	
Total	30	46	49	0	0 / 125																																																																																		
Recall	1.0	1.0	1.0																																																																																				
	A	B	C	Error	Rate	Precision																																																																																	
A	35	0	0	0	0 / 35	0.95																																																																																	
B	1	56	2	0.0508	3 / 59	0.97																																																																																	
C	1	2	61	0.0469	3 / 64	0.97																																																																																	
Total	37	58	63	0.0380	6 / 158																																																																																		
Recall	1.0	0.95	0.95																																																																																				
05_07_Mk_m9_Targ3m_F																																																																																							
<div>▼ TRAINING METRICS - CONFUSION MATRIX ROW</div> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>A</td><td>27</td><td>0</td><td>0</td><td>0</td><td>0 / 27</td><td>0.96</td></tr><tr><td>B</td><td>1</td><td>32</td><td>7</td><td>0.2000</td><td>8 / 40</td><td>0.84</td></tr><tr><td>C</td><td>0</td><td>6</td><td>42</td><td>0.1250</td><td>6 / 48</td><td>0.86</td></tr><tr><td>Total</td><td>28</td><td>38</td><td>49</td><td>0.1217</td><td>14 / 115</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>0.80</td><td>0.88</td><td></td><td></td><td></td></tr></table>			A	B	C	Error	Rate	Precision	A	27	0	0	0	0 / 27	0.96	B	1	32	7	0.2000	8 / 40	0.84	C	0	6	42	0.1250	6 / 48	0.86	Total	28	38	49	0.1217	14 / 115		Recall	1.0	0.80	0.88				<div>▼ VALIDATION METRICS - CONFUSION MATRIX ROW</div> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>Error</td><td>Rate</td><td>Precision</td></tr><tr><td>A</td><td>32</td><td>0</td><td>0</td><td>0</td><td>0 / 32</td><td>0.91</td></tr><tr><td>B</td><td>2</td><td>42</td><td>7</td><td>0.1765</td><td>9 / 51</td><td>0.84</td></tr><tr><td>C</td><td>1</td><td>8</td><td>54</td><td>0.1429</td><td>9 / 63</td><td>0.89</td></tr><tr><td>Total</td><td>35</td><td>50</td><td>61</td><td>0.1233</td><td>18 / 146</td><td></td></tr><tr><td>Recall</td><td>1.0</td><td>0.82</td><td>0.86</td><td></td><td></td><td></td></tr></table>			A	B	C	Error	Rate	Precision	A	32	0	0	0	0 / 32	0.91	B	2	42	7	0.1765	9 / 51	0.84	C	1	8	54	0.1429	9 / 63	0.89	Total	35	50	61	0.1233	18 / 146		Recall	1.0	0.82	0.86			
	A	B	C	Error	Rate	Precision																																																																																	
A	27	0	0	0	0 / 27	0.96																																																																																	
B	1	32	7	0.2000	8 / 40	0.84																																																																																	
C	0	6	42	0.1250	6 / 48	0.86																																																																																	
Total	28	38	49	0.1217	14 / 115																																																																																		
Recall	1.0	0.80	0.88																																																																																				
	A	B	C	Error	Rate	Precision																																																																																	
A	32	0	0	0	0 / 32	0.91																																																																																	
B	2	42	7	0.1765	9 / 51	0.84																																																																																	
C	1	8	54	0.1429	9 / 63	0.89																																																																																	
Total	35	50	61	0.1233	18 / 146																																																																																		
Recall	1.0	0.82	0.86																																																																																				

Таблица 7. Оценка ошибок моделирования (продолжение).

Оценка на обучающем наборе							Оценка на полном наборе						
05_08_Mk_m9_Targ3m_S													
▼ TRAINING METRICS - CONFUSION MATRIX ROW 1							▼ VALIDATION METRICS - CONFUSION MATRIX ROW 1						
	A	B	C	Error	Rate	Precision		A	B	C	Error	Rate	Precision
A	27	0	0	0	0 / 27	0.96	A	32	0	0	0	0 / 32	0.91
B	1	32	7	0.2000	8 / 40	0.82	B	2	40	9	0.2157	11 / 51	0.82
C	0	7	41	0.1458	7 / 48	0.85	C	1	9	53	0.1587	10 / 63	0.85
Total	28	39	48	0.1304	15 / 115		Total	35	49	62	0.1438	21 / 146	
Recall	1.0	0.80	0.85				Recall	1.0	0.78	0.84			
05_09_Rr_m4_Targ3r_F													
HET							HET						
05_10_Rr_m4_Targ3r_S													
HET							HET						
05_11_Rr_m7_Targ3r_F													
HET							HET						
05_12_Rr_m7_Targ3r_S													
HET							HET						

4.15.4 ROC – кривая

Таблица 8. Сравнение ROC – кривых моделирования.

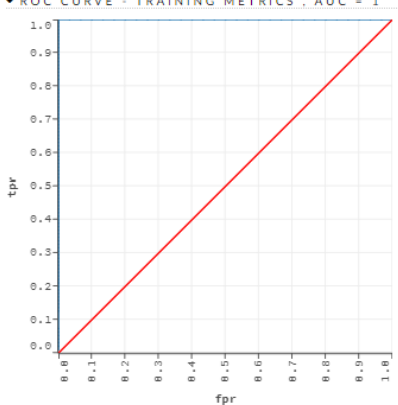
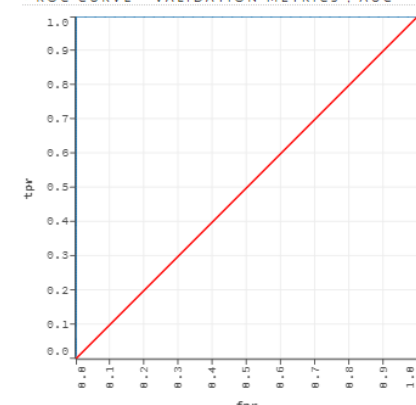
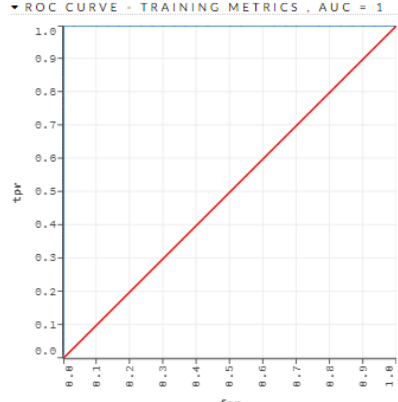
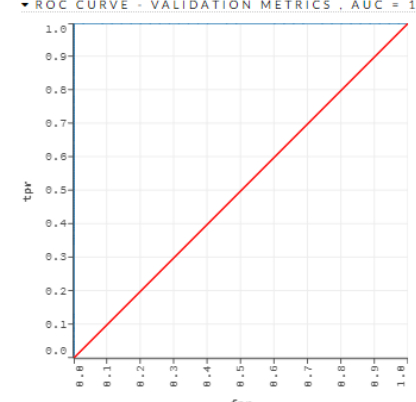
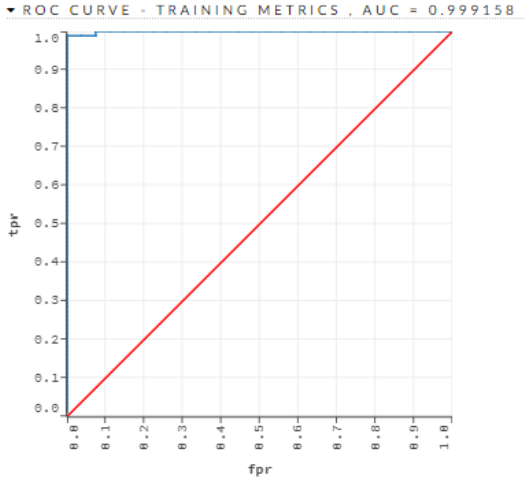
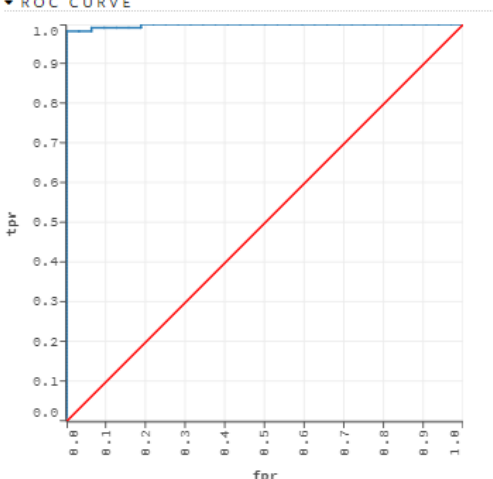
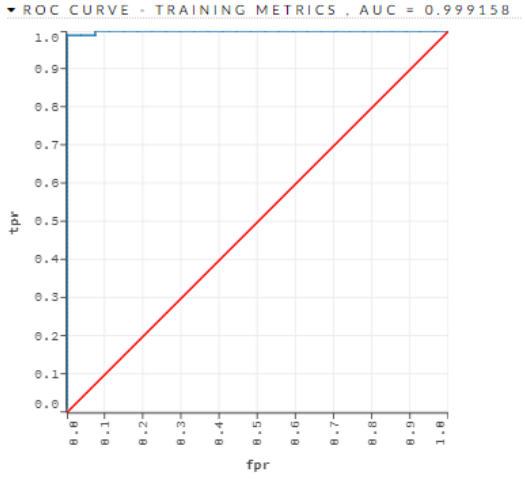
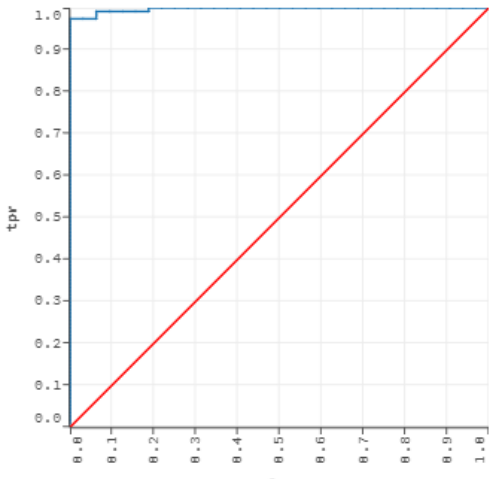
Оценка на обучающем наборе		Оценка на полном наборе	
05_01_Bk_m1_MTarg_F			
<p>▼ ROC CURVE - TRAINING METRICS , AUC = 1</p> 		<p>▼ ROC CURVE - VALIDATION METRICS , AUC = 1</p> 	
05_02_Bk_m1_MTarg_S			
<p>▼ ROC CURVE - TRAINING METRICS , AUC = 1</p> 		<p>▼ ROC CURVE - VALIDATION METRICS , AUC = 1</p> 	

Таблица 8. Сравнение ROC – кривых моделирования (продолжение).

<i>Оценка на обучающем наборе</i>	<i>Оценка на полном наборе</i>
05_03_Bk_m8_MTarg_F	
	
05_04_Bk_m8_MTarg_S	
	
05_05_Mk_m1_Targ3m_F	
нет	нет
05_06_Mk_m1_Targ3m_S	
нет	нет
05_07_Mk_m9_Targ3m_F	
нет	нет
05_08_Mk_m9_Targ3m_S	
нет	нет
05_09_Rr_m4_Targ3r_F	
нет	нет
05_10_Rr_m4_Targ3r_S	
нет	нет
05_11_Rr_m7_Targ3r_F	
нет	нет
05_12_Rr_m7_Targ3r_S	
нет	нет

4.15.5 Метрики модели

Таблица 9. Сравнение метрик моделей.

Оценка на обучающем наборе	Оценка на полном наборе
05_01_Bk_m1_MTarg_F	
▼ OUTPUT - TRAINING METRICS <pre> model a427e15a-1dd7-4828-b409-be3fa87b6a87 model_checksum 4675357570860780905 frame 05_01_Bk_MTarg_m1_F_0.750 frame_checksum -8499038547437041722 description . model_category Binomial scoring_time 1616012841969 predictions . MSE 0.000069 RMSE 0.008286 nobs 125 custom_metric_name . custom_metric_value 0 r2 0.999624 logloss 0.005695 AUC 1 pr_auc 1 Gini 1 mean_per_class_error 0 </pre>	▼ OUTPUT - VALIDATION METRICS <pre> model a427e15a-1dd7-4828-b409-be3fa87b6a87 model_checksum 4675357570860780905 frame Key_Frame__1_RR_v2.hex frame_checksum 3609225813034810295 description . model_category Binomial scoring_time 1616012841977 predictions . MSE 0.010497 RMSE 0.102457 nobs 158 custom_metric_name . custom_metric_value 0 r2 0.939127 logloss 0.031177 AUC 1 pr_auc 1 Gini 1 mean_per_class_error 0 </pre>
05_02_Bk_m1_MTarg_S	
▼ OUTPUT - TRAINING METRICS <pre> model fba73bb7-d8af-4684-a257-1d7e2dbb70c8 model_checksum -6418063867740376216 frame 05_02_Bk_m1_MTarg_S_0.750 frame_checksum -8499038547437041722 description . model_category Binomial scoring_time 1616013337089 predictions . MSE 0.000079 RMSE 0.008888 nobs 125 custom_metric_name . custom_metric_value 0 r2 0.999567 logloss 0.005792 AUC 1 pr_auc 1 Gini 1 mean_per_class_error 0 </pre>	▼ OUTPUT - VALIDATION METRICS <pre> model fba73bb7-d8af-4684-a257-1d7e2dbb70c8 model_checksum -6418063867740376216 frame Key_Frame__1_RR_v2.hex frame_checksum 3609225813034810295 description . model_category Binomial scoring_time 1616013337102 predictions . MSE 0.013646 RMSE 0.116817 nobs 158 custom_metric_name . custom_metric_value 0 r2 0.920868 logloss 0.040903 AUC 1 pr_auc 1 Gini 1 mean_per_class_error 0 </pre>
05_03_Bk_m8_MTarg_F	
▼ OUTPUT - TRAINING METRICS <pre> model c84d1e7f-18b2-4bb6-b999-9149f34e5d62 model_checksum 719885607915930504 frame 05_03_Bk_m8_MTarg_F_0.750 frame_checksum -8499038547437041722 description . model_category Binomial scoring_time 1616013651919 predictions . MSE 0.026210 RMSE 0.161895 nobs 115 custom_metric_name . custom_metric_value 0 r2 0.854113 logloss 0.122035 AUC 0.999158 pr_auc 0.999746 Gini 0.998316 mean_per_class_error 0.005682 residual_deviance 28.068122 null_deviance 125.347919 AIC 72.068122 null_degrees_of_freedom 114 residual_degrees_of_freedom 93 </pre>	▼ OUTPUT - VALIDATION METRICS <pre> model c84d1e7f-18b2-4bb6-b999-9149f34e5d62 model_checksum 719885607915930504 frame Key_Frame__1_RR_v2.hex frame_checksum 3609225813034810295 description . model_category Binomial scoring_time 1616013651967 predictions . MSE 0.026021 RMSE 0.161310 nobs 146 custom_metric_name . custom_metric_value 0 r2 0.847954 logloss 0.119050 AUC 0.997807 pr_auc 0.999406 Gini 0.995614 mean_per_class_error 0.008772 residual_deviance 34.762481 null_deviance 153.753828 AIC 78.762481 null_degrees_of_freedom 145 residual_degrees_of_freedom 124 </pre>

Таблица 9. Сравнение метрик моделей (продолжение).

<i>Оценка на обучающем наборе</i>	<i>Оценка на полном наборе</i>
05_04_Bk_m8_MTarg_S	
<div>▼ OUTPUT - TRAINING METRICS</div> <div> <div>model 72cccf4a-43da-4f63-bc34-2189b9a994c2</div> <div>model_checksum 4571794923473261100</div> <div>frame 05_04_Bk_m8_MTarg_S_0.750</div> <div>frame_checksum -8499038547437041722</div> <div>description .</div> <div>model_category Binomial</div> <div>scoring_time 1616014716159</div> <div>predictions .</div> <div>MSE 0.026639</div> <div>RMSE 0.163213</div> <div>nobs 115</div> <div>custom_metric_name .</div> <div>custom_metric_value 0</div> <div>r2 0.851727</div> <div>logloss 0.123115</div> <div>AUC 0.999158</div> <div>pr_auc 0.999746</div> <div>Gini 0.998316</div> <div>mean_per_class_error 0.005682</div> <div>residual_deviance 28.316344</div> <div>null_deviance 125.347919</div> <div>AIC 60.316344</div> <div>null_degrees_of_freedom 114</div> <div>residual_degrees_of_freedom 99</div> </div>	<div>▼ OUTPUT - VALIDATION METRICS</div> <div> <div>model 72cccf4a-43da-4f63-bc34-2189b9a994c2</div> <div>model_checksum 4571794923473261100</div> <div>frame Key_Frame__1_RR_v2.hex</div> <div>frame_checksum 3609225813034810295</div> <div>description .</div> <div>model_category Binomial</div> <div>scoring_time 1616014716203</div> <div>predictions .</div> <div>MSE 0.026837</div> <div>RMSE 0.163821</div> <div>nobs 146</div> <div>custom_metric_name .</div> <div>custom_metric_value 0</div> <div>r2 0.843185</div> <div>logloss 0.120987</div> <div>AUC 0.997259</div> <div>pr_auc 0.999252</div> <div>Gini 0.994518</div> <div>mean_per_class_error 0.035636</div> <div>residual_deviance 35.328198</div> <div>null_deviance 153.753828</div> <div>AIC 67.328198</div> <div>null_degrees_of_freedom 145</div> <div>residual_degrees_of_freedom 130</div> </div>
05_05_Mk_m1_Targ3m_F	
<div>▼ OUTPUT - TRAINING METRICS</div> <div> <div>model 3230af36-4864-437c-a41d-283191959347</div> <div>model_checksum -2422654752514827456</div> <div>frame 05_05_Mk_m1_Targ3m_F_0.750</div> <div>frame_checksum -8499038547437041722</div> <div>description .</div> <div>model_category Multinomial</div> <div>scoring_time 1616015874035</div> <div>predictions .</div> <div>MSE 0.000351</div> <div>RMSE 0.018740</div> <div>nobs 125</div> <div>custom_metric_name .</div> <div>custom_metric_value 0</div> <div>r2 0.999423</div> <div>logloss 0.015443</div> <div>mean_per_class_error 0</div> <div>AUC NaN</div> <div>pr_auc NaN</div> <div>multinomial_auc_table .</div> <div>multinomial_aucpr_table .</div> </div>	<div>▼ OUTPUT - VALIDATION METRICS</div> <div> <div>model 3230af36-4864-437c-a41d-283191959347</div> <div>model_checksum -2422654752514827456</div> <div>frame Key_Frame__1_RR_v2.hex</div> <div>frame_checksum 3609225813034810295</div> <div>description .</div> <div>model_category Multinomial</div> <div>scoring_time 1616015874036</div> <div>predictions .</div> <div>MSE 0.024655</div> <div>RMSE 0.157020</div> <div>nobs 158</div> <div>custom_metric_name .</div> <div>custom_metric_value 0</div> <div>r2 0.958415</div> <div>logloss 0.094182</div> <div>mean_per_class_error 0.026924</div> <div>AUC NaN</div> <div>pr_auc NaN</div> <div>multinomial_auc_table .</div> <div>multinomial_aucpr_table .</div> </div>
05_06_Mk_m1_Targ3m_S	
<div>▼ OUTPUT - TRAINING METRICS</div> <div> <div>model 2f140483-71d8-460b-b36a-3ff1d48c8363</div> <div>model_checksum 7207282506279251456</div> <div>frame 05_06_Mk_m1_Targ3m_S_0.750</div> <div>frame_checksum -8499038547437041722</div> <div>description .</div> <div>model_category Multinomial</div> <div>scoring_time 1616017487087</div> <div>predictions .</div> <div>MSE 0.001513</div> <div>RMSE 0.038901</div> <div>nobs 125</div> <div>custom_metric_name .</div> <div>custom_metric_value 0</div> <div>r2 0.997515</div> <div>logloss 0.030324</div> <div>mean_per_class_error 0</div> <div>AUC NaN</div> <div>pr_auc NaN</div> <div>multinomial_auc_table .</div> <div>multinomial_aucpr_table .</div> </div>	<div>▼ OUTPUT - VALIDATION METRICS</div> <div> <div>model 2f140483-71d8-460b-b36a-3ff1d48c8363</div> <div>model_checksum 7207282506279251456</div> <div>frame Key_Frame__1_RR_v2.hex</div> <div>frame_checksum 3609225813034810295</div> <div>description .</div> <div>model_category Multinomial</div> <div>scoring_time 1616017487087</div> <div>predictions .</div> <div>MSE 0.029854</div> <div>RMSE 0.172783</div> <div>nobs 158</div> <div>custom_metric_name .</div> <div>custom_metric_value 0</div> <div>r2 0.949647</div> <div>logloss 0.107703</div> <div>mean_per_class_error 0.032574</div> <div>AUC NaN</div> <div>pr_auc NaN</div> <div>multinomial_auc_table .</div> <div>multinomial_aucpr_table .</div> </div>

Таблица 9. Сравнение метрик моделей (продолжение).

Оценка на обучающем наборе		Оценка на полном наборе	
05_07_Mk_m9_Targ3m_F			
▼ OUTPUT - TRAINING METRICS		▼ OUTPUT - VALIDATION METRICS	
model	bb373d1e-e658-410d-9836-60e434c6764c	model	bb373d1e-e658-410d-9836-60e434c6764c
model_checksum	-4117968269583996000	model_checksum	-4117968269583996000
frame	05_07_Mk_m9_Targ3m_F_0.750	frame	Key_Frame__1_RR_v2.hex
frame_checksum	-8499038547437041722	frame_checksum	3609225813034810295
description	.	description	.
model_category	Multinomial	model_category	Multinomial
scoring_time	1616018123959	scoring_time	1616018123971
predictions	.	predictions	.
MSE	0.140112	MSE	0.138441
RMSE	0.374315	RMSE	0.372077
nobs	115	nobs	146
custom_metric_name	.	custom_metric_name	.
custom_metric_value	0	custom_metric_value	0
r2	0.773585	r2	0.771398
logloss	0.441767	logloss	0.435727
mean_per_class_error	0.108333	mean_per_class_error	0.106443
AUC	NaN	AUC	NaN
pr_auc	NaN	pr_auc	NaN
multinomial_auc_table	.	multinomial_auc_table	.
multinomial_aucpr_table	.	multinomial_aucpr_table	.
residual_deviance	101.606515	residual_deviance	127.232219
null_deviance	246.941441	null_deviance	311.300199
AIC	NaN	AIC	NaN
null_degrees_of_freedom	114	null_degrees_of_freedom	145
residual_degrees_of_freedom	71	residual_degrees_of_freedom	102

05_08_Mk_m9_Targ3m_S			
▼ OUTPUT - TRAINING METRICS		▼ OUTPUT - VALIDATION METRICS	
model	e3f51438-11bb-4992-a7b2-847d75b2f410	model	e3f51438-11bb-4992-a7b2-847d75b2f410
model_checksum	-9010798361369360592	model_checksum	-9010798361369360592
frame	05_08_Mk_m9_Targ3m_S_0.750	frame	Key_Frame__1_RR_v2.hex
frame_checksum	-8499038547437041722	frame_checksum	3609225813034810295
description	.	description	.
model_category	Multinomial	model_category	Multinomial
scoring_time	1616020328939	scoring_time	1616020328967
predictions	.	predictions	.
MSE	0.147234	MSE	0.145085
RMSE	0.383710	RMSE	0.380900
nobs	115	nobs	146
custom_metric_name	.	custom_metric_name	.
custom_metric_value	0	custom_metric_value	0
r2	0.762077	r2	0.760428
logloss	0.459992	logloss	0.452312
mean_per_class_error	0.115278	mean_per_class_error	0.124805
AUC	NaN	AUC	NaN
pr_auc	NaN	pr_auc	NaN
multinomial_auc_table	.	multinomial_auc_table	.
multinomial_aucpr_table	.	multinomial_aucpr_table	.
residual_deviance	105.798210	residual_deviance	132.075060
null_deviance	246.941441	null_deviance	311.300199
AIC	NaN	AIC	NaN
null_degrees_of_freedom	114	null_degrees_of_freedom	145
residual_degrees_of_freedom	89	residual_degrees_of_freedom	120
05_09_Rr_m4_Targ3r_F			
▼ OUTPUT - TRAINING METRICS		▼ OUTPUT - VALIDATION METRICS	
model	e4c558a7-e30c-4aec-a858-5fa0e29a9429	model	e4c558a7-e30c-4aec-a858-5fa0e29a9429
model_checksum	-439472540292352192	model_checksum	-439472540292352192
frame	05_09_Rr_m4_Targ3r_F_0.750	frame	Key_Frame__1_RR_v2.hex
frame_checksum	-8499038547437041722	frame_checksum	3609225813034810295
description	.	description	.
model_category	Regression	model_category	Regression
scoring_time	1616021114045	scoring_time	1616021114057
predictions	.	predictions	.
MSE	0.159760	MSE	0.164396
RMSE	0.399699	RMSE	0.405458
nobs	115	nobs	146
custom_metric_name	.	custom_metric_name	.
custom_metric_value	0	custom_metric_value	0
r2	0.741835	r2	0.728540
mean_residual_deviance	0.159760	mean_residual_deviance	0.164396
mae	0.317318	mae	0.315061
rmsle	0.130890	rmsle	0.129972
residual_deviance	18.372355	residual_deviance	24.001872
null_deviance	71.165217	null_deviance	88.546767
AIC	149.436080	AIC	184.730735
null_degrees_of_freedom	114	null_degrees_of_freedom	145
residual_degrees_of_freedom	99	residual_degrees_of_freedom	130

Таблица 9. Сравнение метрик моделей (продолжение).

Оценка на обучающем наборе		Оценка на полном наборе	
05_10_Rr_m4_Targ3r_S			
▼ OUTPUT - TRAINING METRICS		▼ OUTPUT - VALIDATION METRICS	
<div>model4064059d-cd68-414d-a849-9ea65a11377c</div> <div>model_checksum5702330899411735360</div> <div>frame05_10_Rr_m4_Targ3r_S_0.750</div> <div>frame_checksum-8499038547437041722</div> <div>description.</div> <div>model_categoryRegression</div> <div>scoring_time1616021983753</div> <div>predictions.</div> <div>MSE0.161069</div> <div>RMSE0.401334</div> <div>nobs115</div> <div>custom_metric_name.</div> <div>custom_metric_value0</div> <div>r20.739719</div> <div>mean_residual_deviance0.161069</div> <div>mae0.318552</div> <div>rmsle0.131300</div> <div>residual_deviance18.522945</div> <div>null_deviance71.165217</div> <div>AIC142.374848</div> <div>null_degrees_of_freedom114</div> <div>residual_degrees_of_freedom103</div>		<div>model4064059d-cd68-414d-a849-9ea65a11377c</div> <div>model_checksum5702330899411735360</div> <div>frameKey_Frame__1_RR_v2.hex</div> <div>frame_checksum3609225813034810295</div> <div>description.</div> <div>model_categoryRegression</div> <div>scoring_time1616021983760</div> <div>predictions.</div> <div>MSE0.166716</div> <div>RMSE0.408309</div> <div>nobs146</div> <div>custom_metric_name.</div> <div>custom_metric_value0</div> <div>r20.724710</div> <div>mean_residual_deviance0.166716</div> <div>mae0.317564</div> <div>rmsle0.130781</div> <div>residual_deviance24.340545</div> <div>null_deviance88.546767</div> <div>AIC178.776432</div> <div>null_degrees_of_freedom145</div> <div>residual_degrees_of_freedom134</div>	
05_11_Rr_m7_Targ3r_F			
▼ OUTPUT - TRAINING METRICS		▼ OUTPUT - VALIDATION METRICS	
<div>modele36c6dd6-20de-4640-b993-263841216a34</div> <div>model_checksum-7298735667466584432</div> <div>frame05_11_Rr_m7_Targ3r_F_0.750</div> <div>frame_checksum-8499038547437041722</div> <div>description.</div> <div>model_categoryRegression</div> <div>scoring_time1616023014234</div> <div>predictions.</div> <div>MSE0.207822</div> <div>RMSE0.455875</div> <div>nobs115</div> <div>custom_metric_name.</div> <div>custom_metric_value0</div> <div>r20.664168</div> <div>mean_residual_deviance0.052172</div> <div>mae0.365604</div> <div>rmsle0.146470</div> <div>residual_deviance5.999763</div> <div>null_deviance18.601232</div> <div>AICNaN</div> <div>null_degrees_of_freedom114</div> <div>residual_degrees_of_freedom102</div>		<div>modele36c6dd6-20de-4640-b993-263841216a34</div> <div>model_checksum-7298735667466584432</div> <div>frameKey_Frame__1_RR_v2.hex</div> <div>frame_checksum3609225813034810295</div> <div>description.</div> <div>model_categoryRegression</div> <div>scoring_time1616023014243</div> <div>predictions.</div> <div>MSE0.219778</div> <div>RMSE0.468805</div> <div>nobs146</div> <div>custom_metric_name.</div> <div>custom_metric_value0</div> <div>r20.637091</div> <div>mean_residual_deviance0.053392</div> <div>mae0.371596</div> <div>rmsle0.148539</div> <div>residual_deviance7.795210</div> <div>null_deviance22.762012</div> <div>AICNaN</div> <div>null_degrees_of_freedom145</div> <div>residual_degrees_of_freedom133</div>	
05_12_Rr_m7_Targ3r_S			
▼ OUTPUT - TRAINING METRICS		▼ OUTPUT - VALIDATION METRICS	
<div>model3ac922a4-4679-4de3-a12e-a7fc22031b0c</div> <div>model_checksum-5517217425360759936</div> <div>frame05_12_Rr_m7_Targ3r_S_0.750</div> <div>frame_checksum-8499038547437041722</div> <div>description.</div> <div>model_categoryRegression</div> <div>scoring_time1616024334028</div> <div>predictions.</div> <div>MSE0.207888</div> <div>RMSE0.455948</div> <div>nobs115</div> <div>custom_metric_name.</div> <div>custom_metric_value0</div> <div>r20.664061</div> <div>mean_residual_deviance0.052219</div> <div>mae0.365741</div> <div>rmsle0.146513</div> <div>residual_deviance6.005202</div> <div>null_deviance18.601232</div> <div>AICNaN</div> <div>null_degrees_of_freedom114</div> <div>residual_degrees_of_freedom104</div>		<div>model3ac922a4-4679-4de3-a12e-a7fc22031b0c</div> <div>model_checksum-5517217425360759936</div> <div>frameKey_Frame__1_RR_v2.hex</div> <div>frame_checksum3609225813034810295</div> <div>description.</div> <div>model_categoryRegression</div> <div>scoring_time1616024334051</div> <div>predictions.</div> <div>MSE0.219812</div> <div>RMSE0.468841</div> <div>nobs146</div> <div>custom_metric_name.</div> <div>custom_metric_value0</div> <div>r20.637036</div> <div>mean_residual_deviance0.053455</div> <div>mae0.371866</div> <div>rmsle0.148604</div> <div>residual_deviance7.804500</div> <div>null_deviance22.762012</div> <div>AICNaN</div> <div>null_degrees_of_freedom145</div> <div>residual_degrees_of_freedom135</div>	

4.16 Загрузка моделей в базу знаний и получение их уникальных идентификаторов UUID

Для загрузки моделей в БЗ и их эксплуатации в качестве одного из вариантов используется web-интерфейс инструмента «A swagger API», доступный по ссылке: <https://knowledge-base-2021.herokuapp.com/apidocs/>.

Загрузка осуществляется командой «POST /upload-model» с указанием файла модели в json-формате, предварительно сохраненного средствами системы H2O.

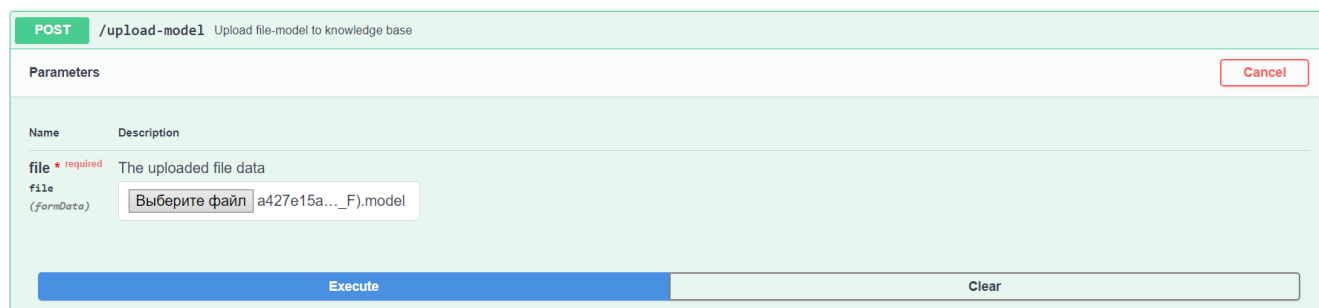
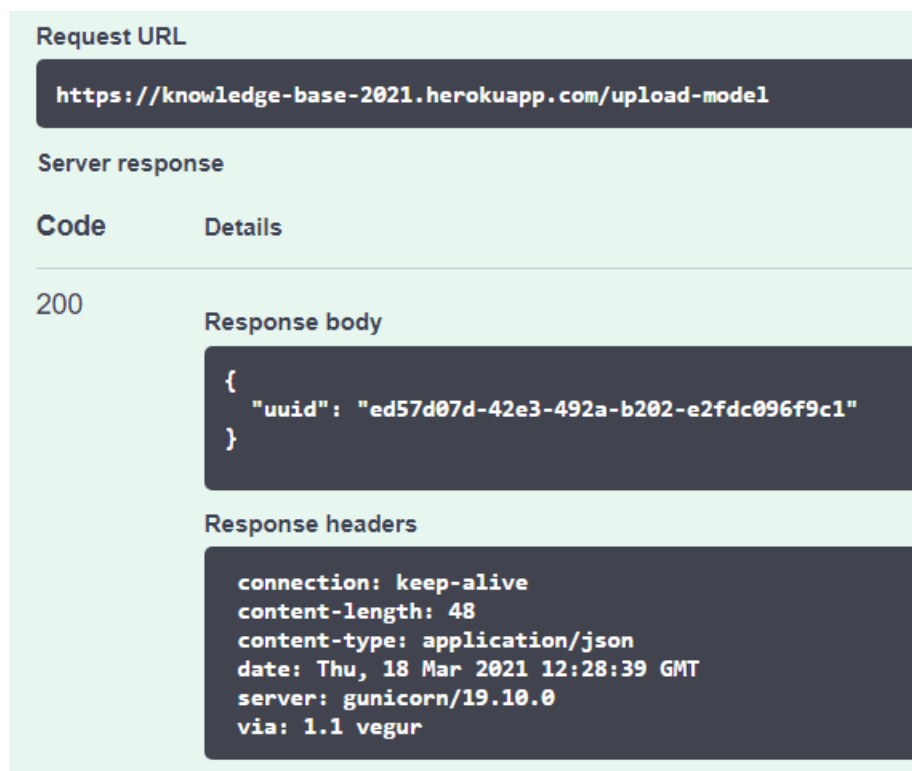


Рис. 40. Интерфейс загрузки в БЗ на примере модели «05_01_Bk_MTarg_m1_F».



На рис. 41 представлен ответ сервера базы знаний, в теле которого содержится уникальный идентификатор модели «**uuid**», который используется для последующих обращений к модели, загруженной и находящейся в этой базе.

Рис. 41. Ответ сервера после отправки файла.

Кроме того, сервер даёт информацию для программного доступа к моделям по API с помощью утилиты «curl» с параметрами:

```
curl -X POST "https://knowledge-base-2021.herokuapp.com/upload-model" -H "accept: application/json" -H "Content-Type: multipart/form-data" -F "file=@a427e15a-1dd7-4828-b409-be3fa87b6a87(05_01_Bk_MTarg_m1_F).model"
```

Выводы по работе

В текущей работе проводилось создание, обучение и тестирование моделей интеллектуального анализа данных и машинного обучения средствами платформы «H2O» с использованием исходных данных медицинского направления, содержащих параметры обследований пациентов на наличие признаков ректального рака.

Результатом работы стали 12 обученных для задач регрессии, бинарной и множественной классификации на полном и усечённом наборах данных моделей, протестированных на полном наборе данных. Все модели были загружены в базу знаний средствами web-интерфейса инструмента «A swagger API».

Анализ результатов работы моделей на основе предложенных исходных данных показывает, что:

1. **Задача бинарной классификации** наилучшим образом была решена методом градиентного бустинга «m1» моделью, обученной на полном наборе данных, где выявился самый высокий уровень корреляции признаков, а также, наблюдались наименьшие среднеквадратичные отклонения и отсутствие ошибок классификации. При этом, точность распознавания у модели с усечённым набором входных параметров оказалась чуть ниже, но сопоставимой с точностью модели, обученной полным набором данных. По мнению автора данной работы, это обусловлено потерей корреляционной связи с отсутствующими параметрами усечённого набора данных. Это подтверждает и пара моделей, работающих по методу «m8» - биномиальной логистической регрессии класса обобщенных линейных моделей, где обученная полным набором данных модель, показала немногим более высокую точность распознавания, но с наличием некоторого числа ошибок распознавания признаков класса «F», что может быть связано с меньшим количеством данных класса «F» при усечении набора и недостаточным уровнем обученности модели на этом классе.
2. **Задача множественной классификации** наиболее успешно была решена методом «m1» - градиентным бустингом с самым высоким значением коэффициента корреляции «r2» у модели, обученной полным набором данных, при этом метод «m9» линейной модели мультиномиальной классификации дал в двое большее среднеквадратическое отклонение у обеих моделей. Точность классификации у метода «m9» значительно ниже, чем у метода «m1», причём тенденция лучшего результата наблюдается у моделей, обученных полным набором данных для обоих методов «m1» и «m9». В данном случае метод «m9» оказался более чувствителен к усечению набора обучающих данных, что отразилось на большем числе ошибок классификации.
3. **Задача регрессии** решена с наилучшими показателями методом Гаусса «m4» обоими моделями, где минимальное преимущество осталось за моделью, обученной полным набором данных. В свою очередь метод «гамма»-регрессии «m7», несколько хуже обнаружил корреляционные зависимости в тестовом наборе данных по сравнению с методом «m4», где полная модель проявила минимальный перевес в точности прогнозирования в отношении модели с усечённым набором данных, при этом, обе модели метода «m7» показали

примерно одинаковые отклонения в прогнозах и почти равные значения коэффициента корреляции.

Несмотря на отсутствие ошибок классификации в соответствующих моделях их точность не равна 1, что может быть связано с пропуском строк, содержащих неполные исходные данные, установленным соответствующей опцией при построении моделей системой «H2O», что даёт более «чистые» результаты анализа при поиске связей между параметрами.

Учитывая несущественно малую разницу в точности результатов моделей с полным и усечённым набором данных относительно всех решаемых в данной работе задач всеми методами, можно гарантировано использовать модели с ограниченным набором данных для практического применения.

Список литературы

Яковина И.Н., Осипенко И.В.

Интеллектуальный анализ данных и машинное обучение. Методические указания к лабораторным работам №1, 2. Электронный документ. НГТУ. АВТФ. 2021

URL:

https://docs.google.com/document/d/1L1YnY3VL4JJe8if1dInXKDGKsFpegasGclP1U_mho_0/edit#