

Анализ риска лесных пожаров на основе данных ДЗЗ и гибридных методов машинного обучения

Ю.В. Пестова (yupest@icc.ru),
м.н.с. ИДСТУ СО РАН, г. Иркутск

О.А. Николайчук, д.т.н., ИДСТУ СО РАН, г. Иркутск

Тема диссертации:

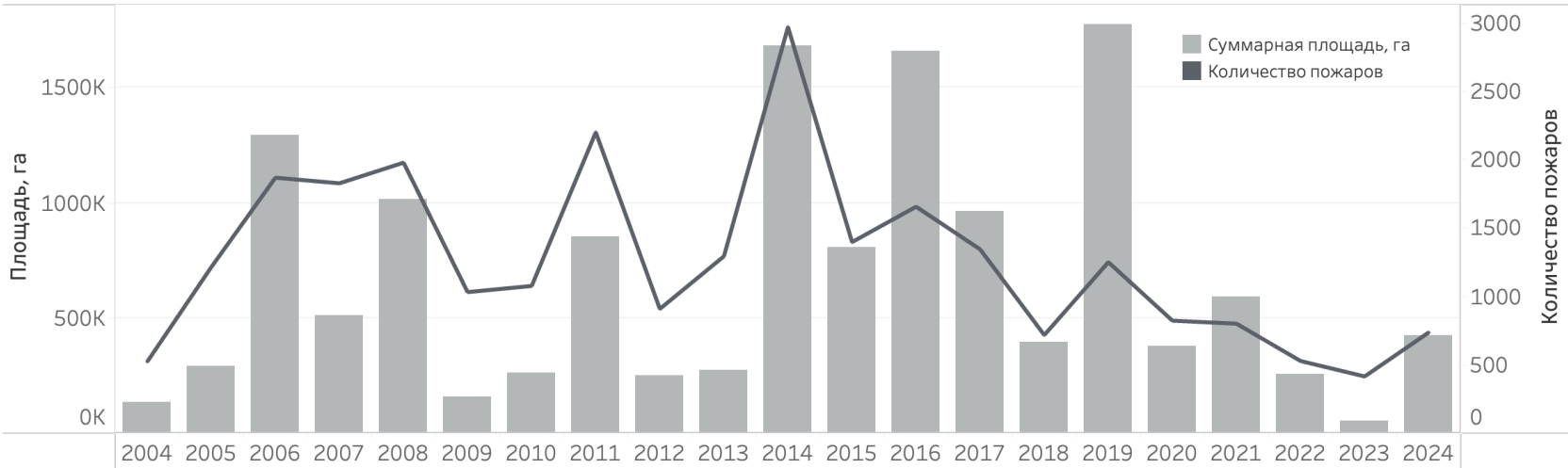
Автоматизированная информационная система прогнозирования риска лесных пожаров территории на основе методов машинного обучения на примере Иркутской области

Описание проблемы. Тенденции лесных пожаров на территории Иркутской области

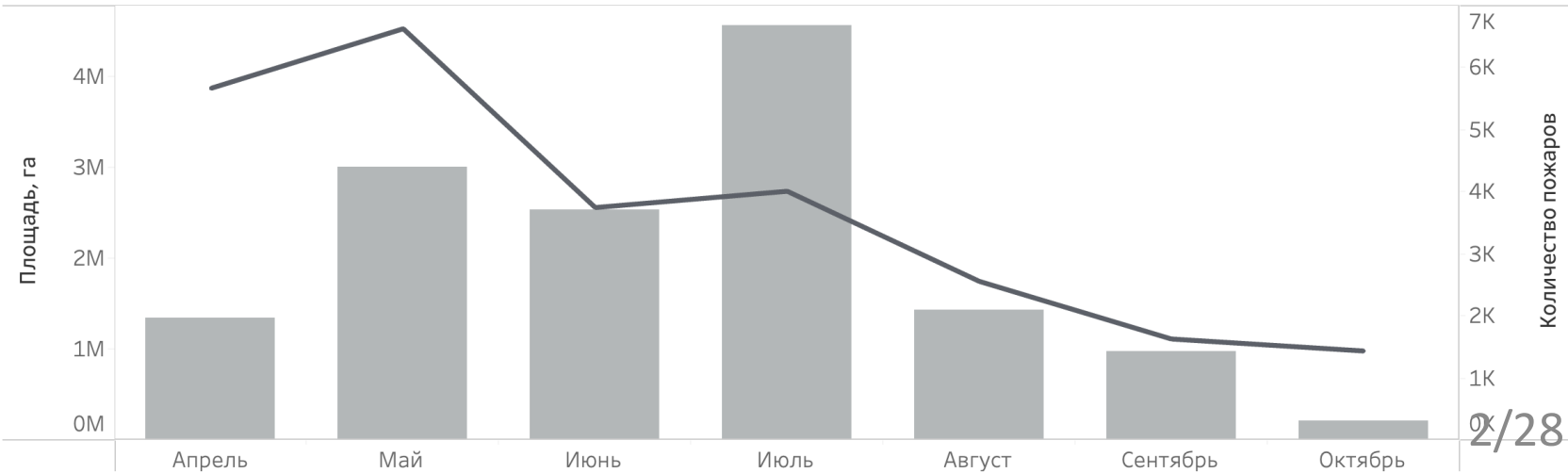
В 2024 г. по данным Рослесхоза площадь, пройденная огнем, составляет 435,3 тыс. га.

На охрану лесов от пожаров в области на 2025 г. запланированы финансовые средства в объеме 2,8 млрд руб.

Тенденции лесных пожаров в Иркутской области



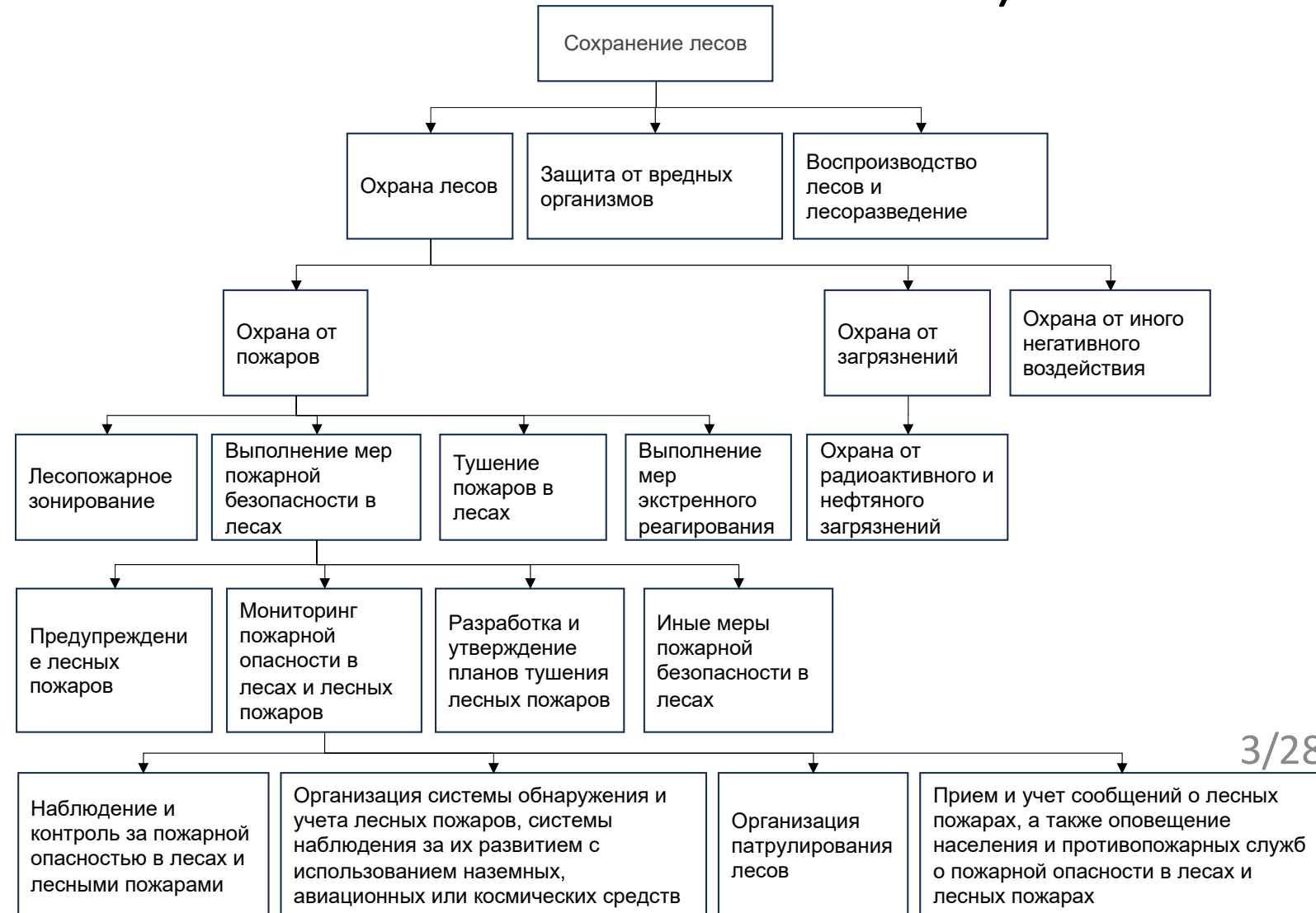
Тенденции пожароопасного периода



Меры пожарной безопасности в лесах (Лесной кодекс от 04.12.2006 N 200-ФЗ)

1. Предупреждение лесных пожаров;
2. Мониторинг пожарной опасности в лесах и лесных пожаров, в том числе с использованием специализированного ПО, позволяющего осуществлять указанную деятельность в автоматическом режиме с использованием ГИС;
3. Разработка и утверждение планов тушения лесных пожаров и др.

Лесной кодекс Российской Федерации от 04.12.2006 N 200-ФЗ (ред. от 26.12.2024). [Электронный ресурс]. URL: <https://normativ.kontur.ru/document?moduleId=1&documentId=484624>



Определения

- **Прогнозирование лесных пожаров¹** – определение вероятности возникновения, разрастания лесных пожаров во времени и пространстве на основе анализа данных мониторинга лесных пожаров.
- **Оценка пожарной опасности в лесах²** выполняется на основе степени пожарной опасности в лесу по условиям погоды, которая определяется по принятому в лесном хозяйстве комплексному показателю В. Г. Нестерова:

$$\text{КПО}_H = \sum_{i=1}^n T_d \cdot (T - T_d),$$

где T – температура воздуха,
 T_d – температура точки росы,
 n – число дней без осадков

Класс пожарной опасности по условиям погоды	Значение комплексного показателя	Степень пожарной опасности
I	до 300	—
II	от 301 до 1000	Малая
III	от 1001 до 4000	Средняя
IV	от 4001 до 10000	Высокая
V	более 10000	Чрезвычайная

¹ Лесной кодекс Российской Федерации от 04.12.2006 N 200-ФЗ (ред. от 26.12.2024). [Электронный ресурс]. URL: <https://normativ.kontur.ru/document?moduleId=1&documentId=484624> (дата обращения: 01.02.2025)

² ГОСТ Р 22.1.09-99. Безопасность в чрезвычайных ситуациях. Приложение А. (дата утверждения: 01.01.2000).

Актуальность исследования (I)

- Существующие программные системы мониторинга лесных пожаров предлагают ограниченный функционал в свободных версиях (оперативное отображение термоточек) и не предусматривают прогнозирование рисков.
- Исследователи отмечают невозможность адаптации существующих моделей машинного обучения к географическим особенностям рассматриваемой территорий.
- Показатель Нестерова, используемый для оценки пожарной опасности, установленный лесохозяйственными регламентами не учитывает рельеф, тип растительности, антропогенную нагрузку и метеорологические характеристики почвы и ветра, что может приводить к неточностям в прогнозах.
- Необходимость сокращения существенных финансовых затрат авиационного мониторинга и обеспечения мониторинга удаленных районов.

Актуальность исследования (II)

Отсутствуют (1) **единые методологические подходы оценки рисков** лесных пожаров территории, (2) **эффективные методы оперативной оценки риска** на заданный период с заданными индивидуальными характеристиками территории с высокой точностью, (3) а также **модели машинного обучения для оценки риска** территории Иркутской области с учетом ее географических особенностей.

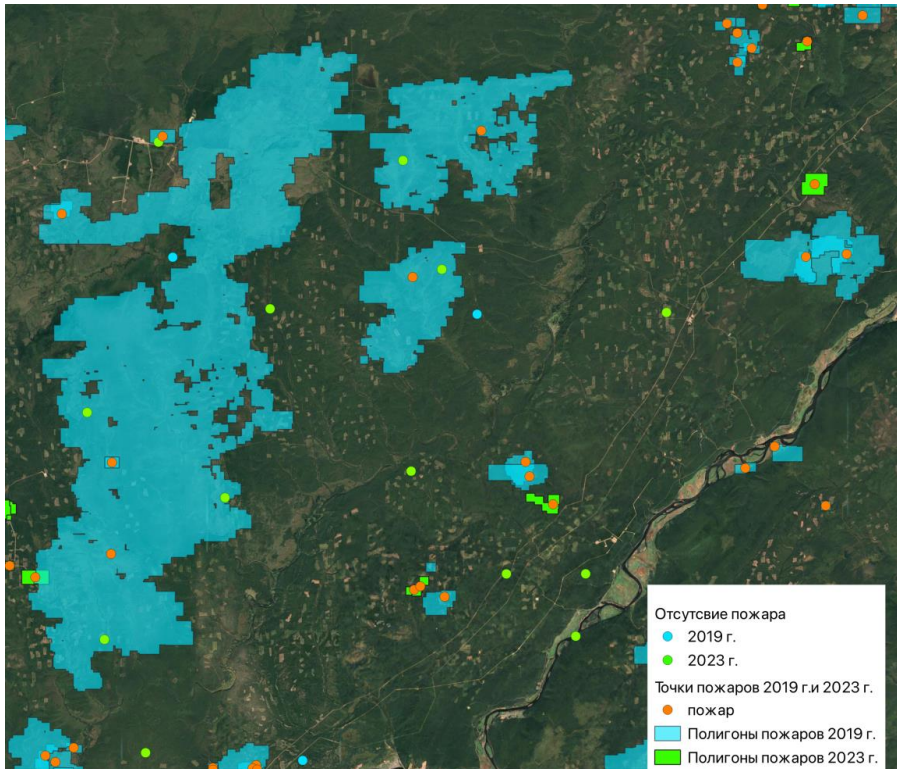
В этой связи разработка АИС прогнозирования рисков на основе машинного обучения, способной анализировать гетерогенные данные (спутниковые снимки, метеоданные, почвенные характеристики, историю пожаров), представляется критически важной для снижения риска и смягчения последствий лесных пожаров.

Концепция информационного процесса

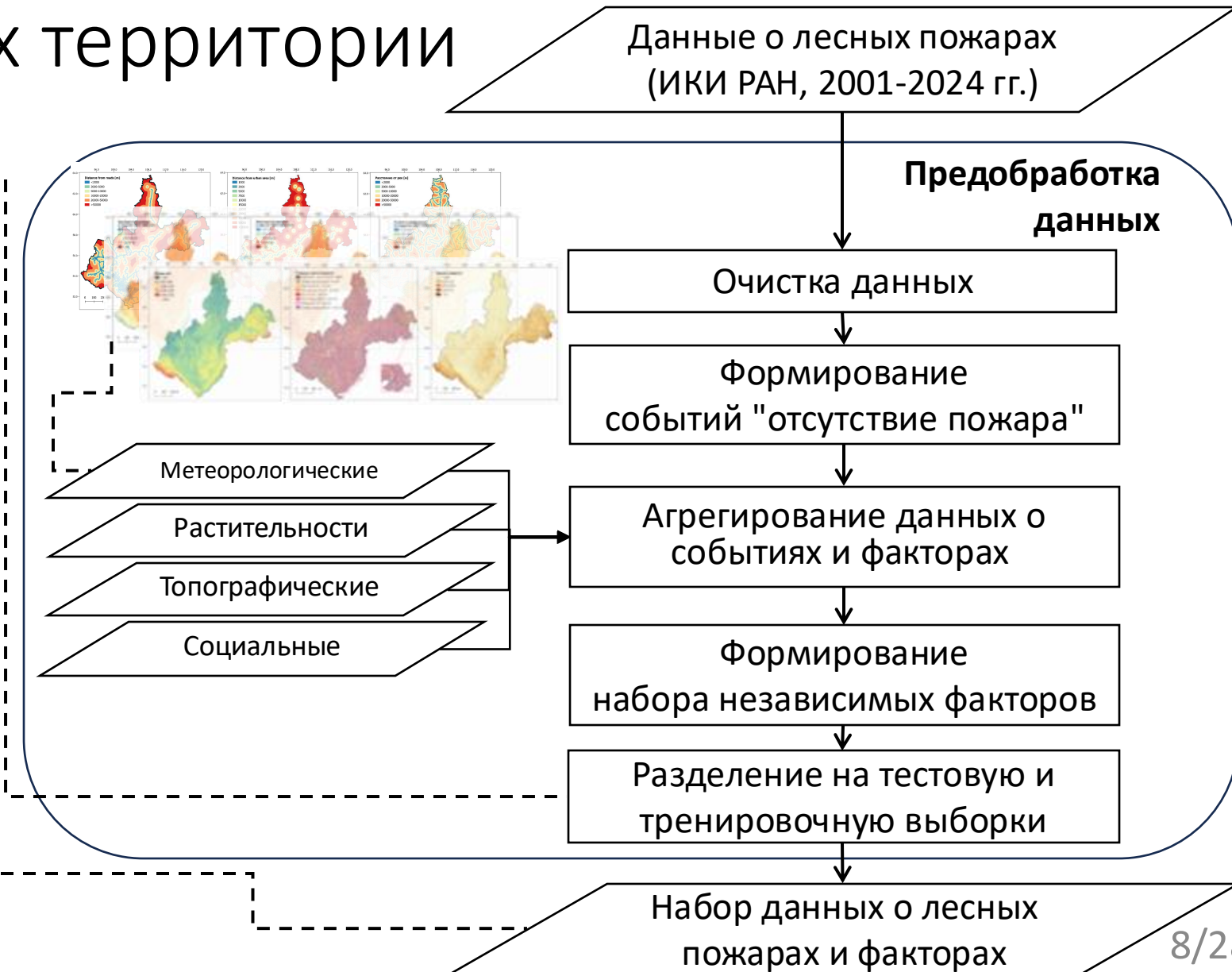


Основные этапы предобработки данных о лесных пожарах и факторах территории

Тестовая выборка – 8746 записей (16%)
за каждые 8 лет (2008, 2016 и 2024 года)
Обучающая выборка – 44394



Исторические пожары (точки «наличие пожара»)
и сгенерированные (точки «отсутствие пожаров»)

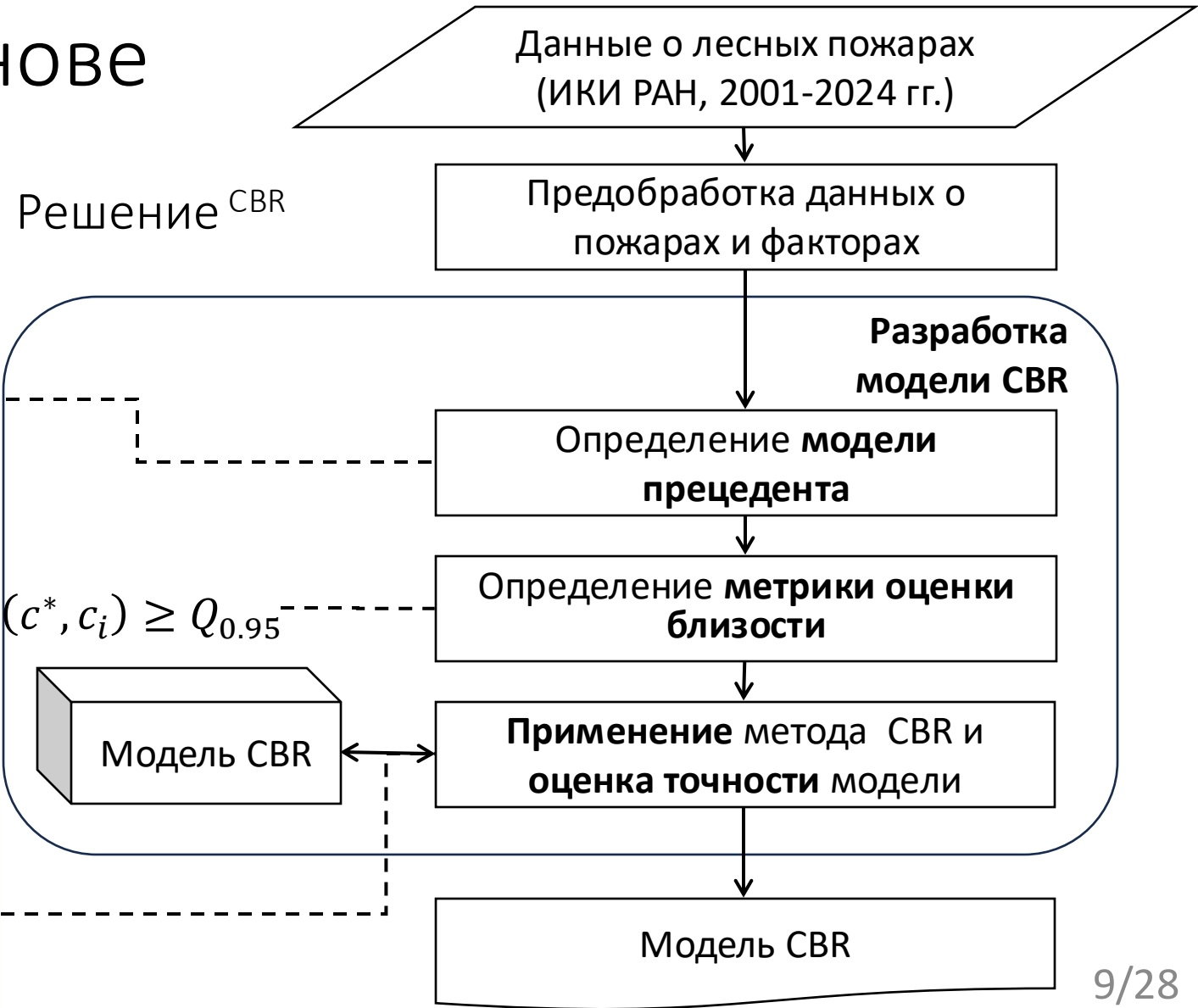
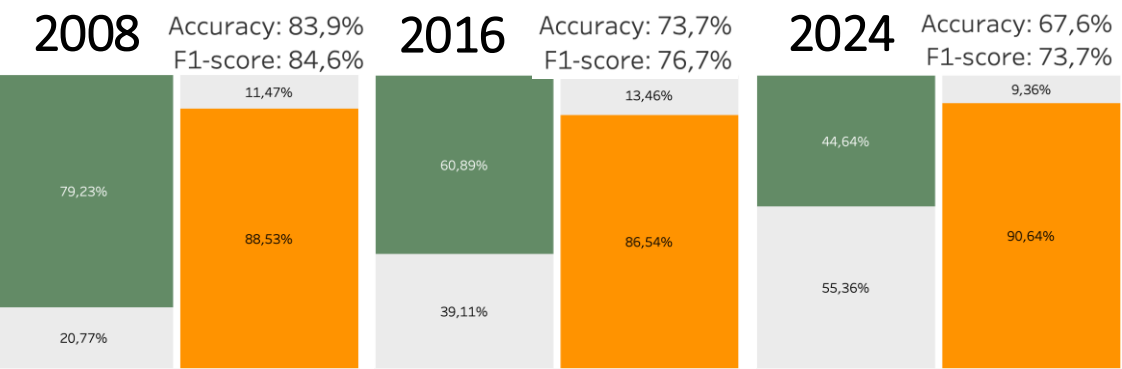


Основные этапы разработки модели прогнозирования риска лесных пожаров на основе CBR

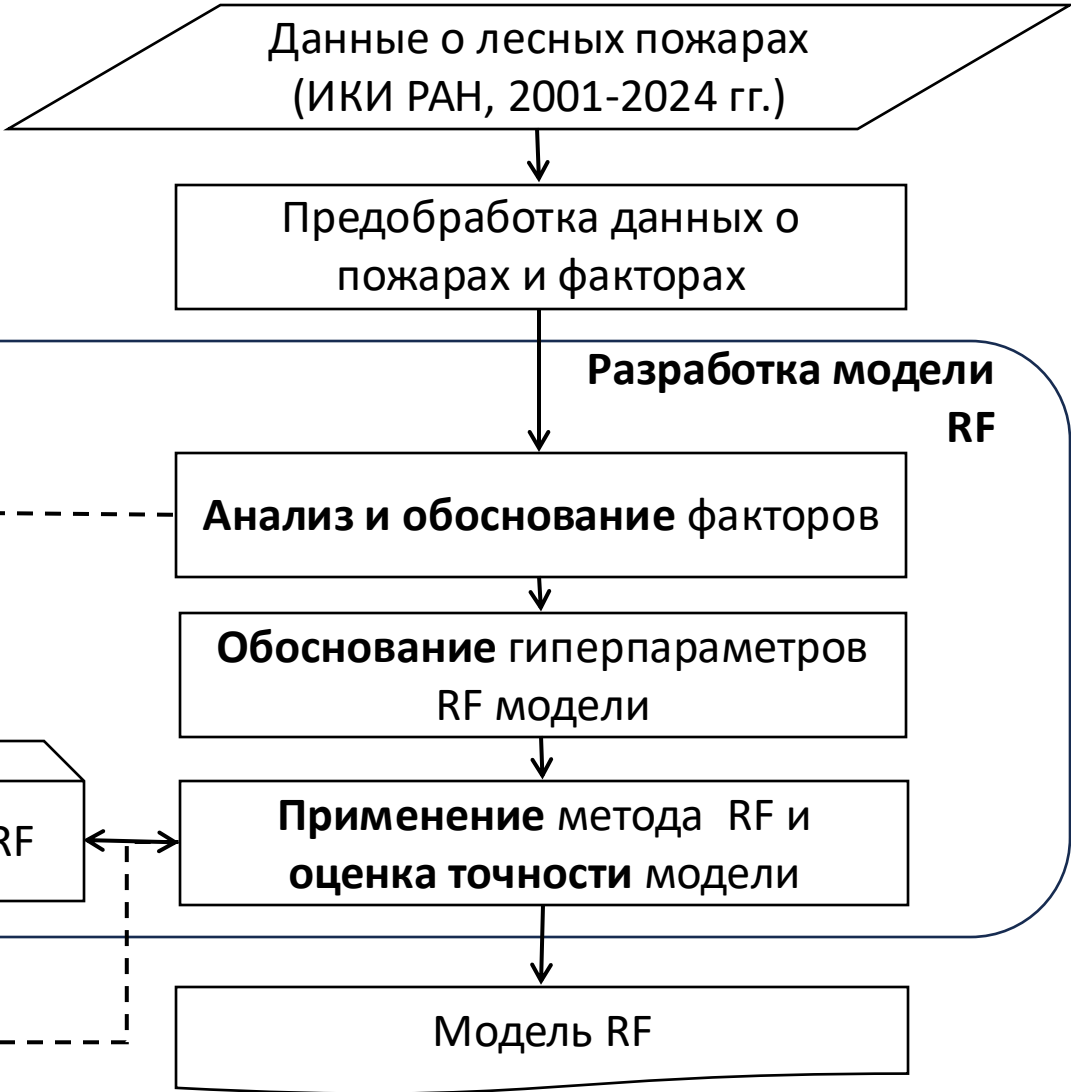
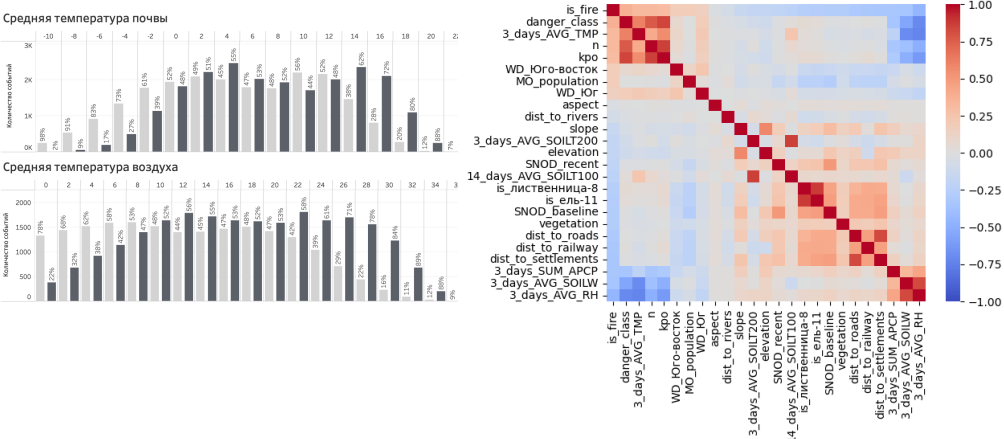
$M_{\text{задача}_{\text{cbr}}}^{\text{задача}} = \text{Проблема}^{\text{CBR}} \rightarrow \text{Решение}^{\text{CBR}}$

- Размер базы прецедентов для тестовых периодов составляет:
- 5435 для 2008 г.,
 - 18296 для 2016 г.,
 - 25833 для 2024 г.

$P(c^*) \in [0; 1] = \frac{1}{N_{Q_{0.95}}} \sum_{i \in N_{Q_{0.95}}} s_i(c^*, c_i), \text{ где } s_i(c^*, c_i) \geq Q_{0.95}$

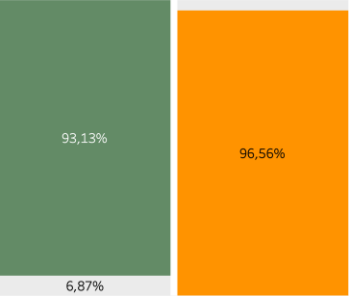


Основные этапы разработки модели прогнозирования риска лесных пожаров на основе Random Forest



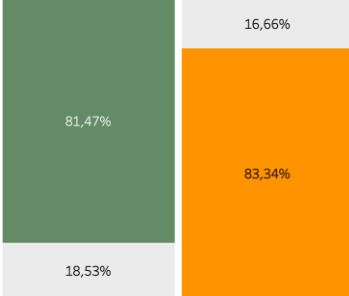
2008

Accuracy: 94,8%
F1-score: 94,9%



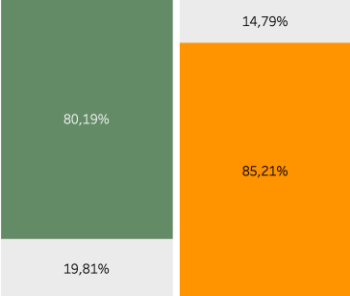
2016

Accuracy: 82,4%
F1-score: 82,6%

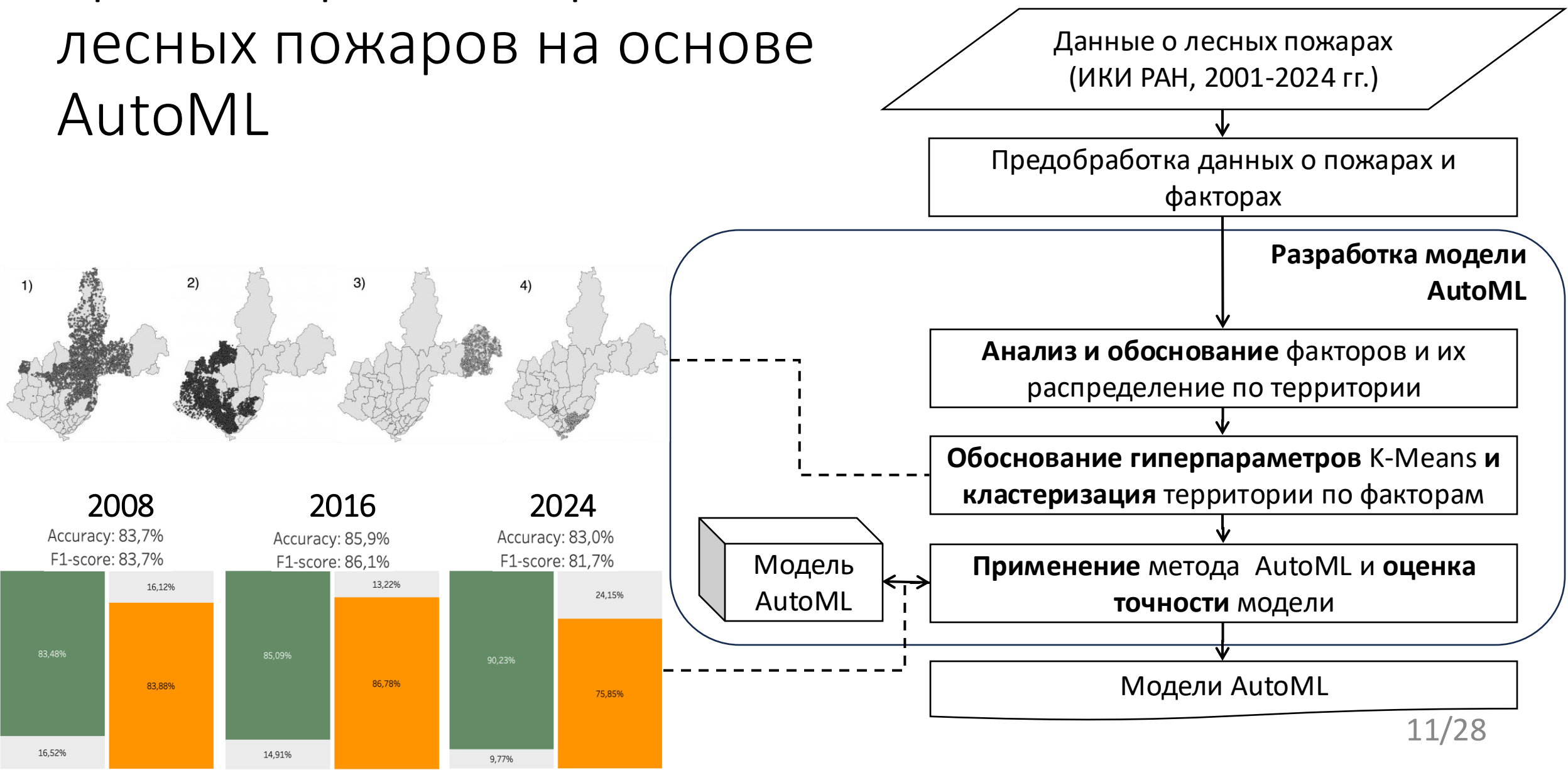


2024

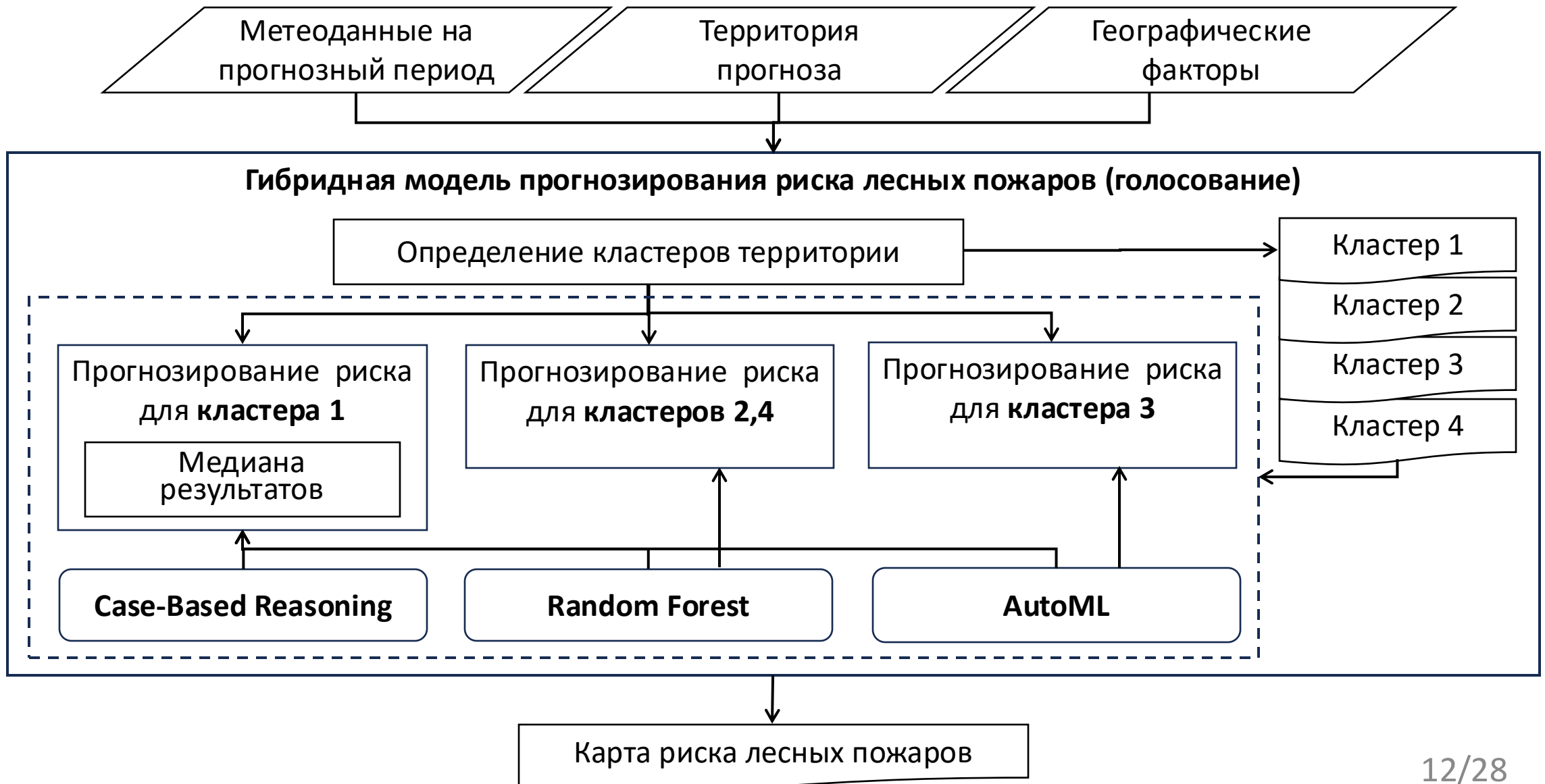
Accuracy: 82,7%
F1-score: 83,1%



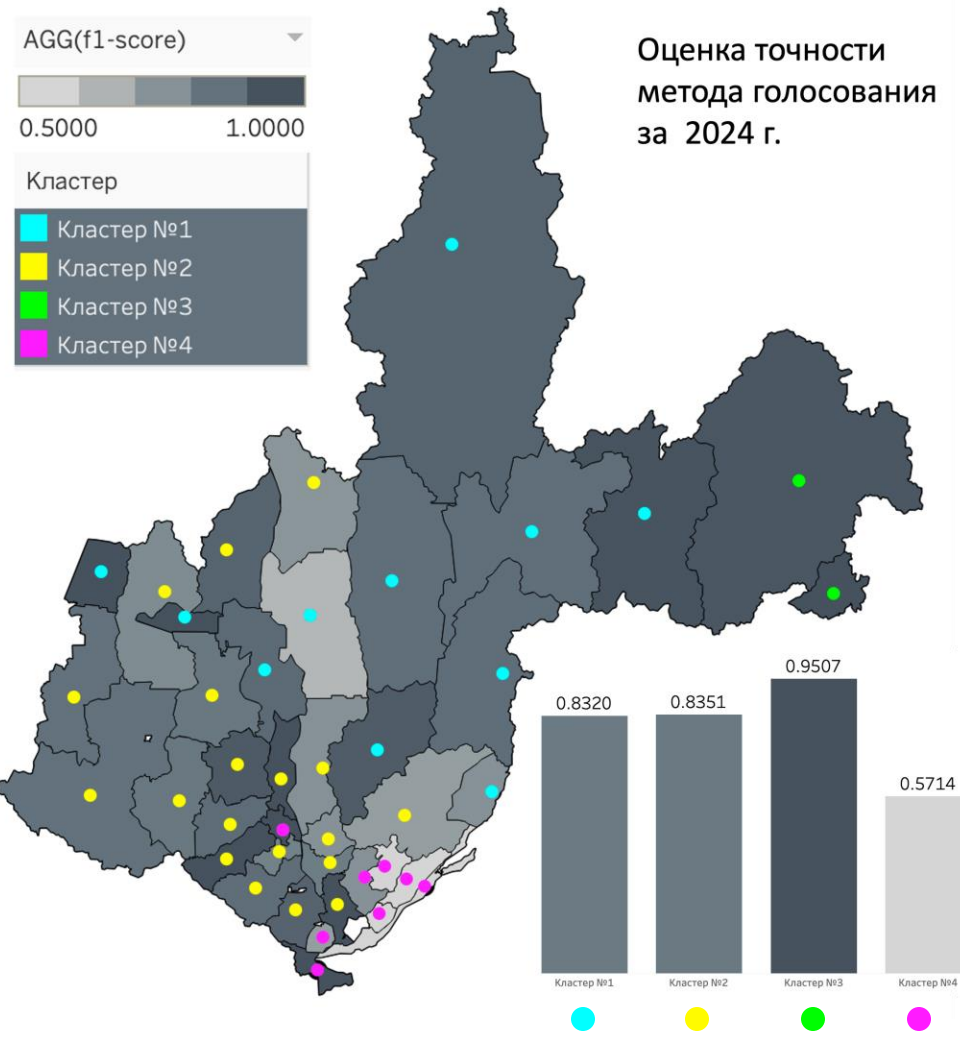
Основные этапы разработки модели прогнозирования риска лесных пожаров на основе AutoML



Гибридная модель

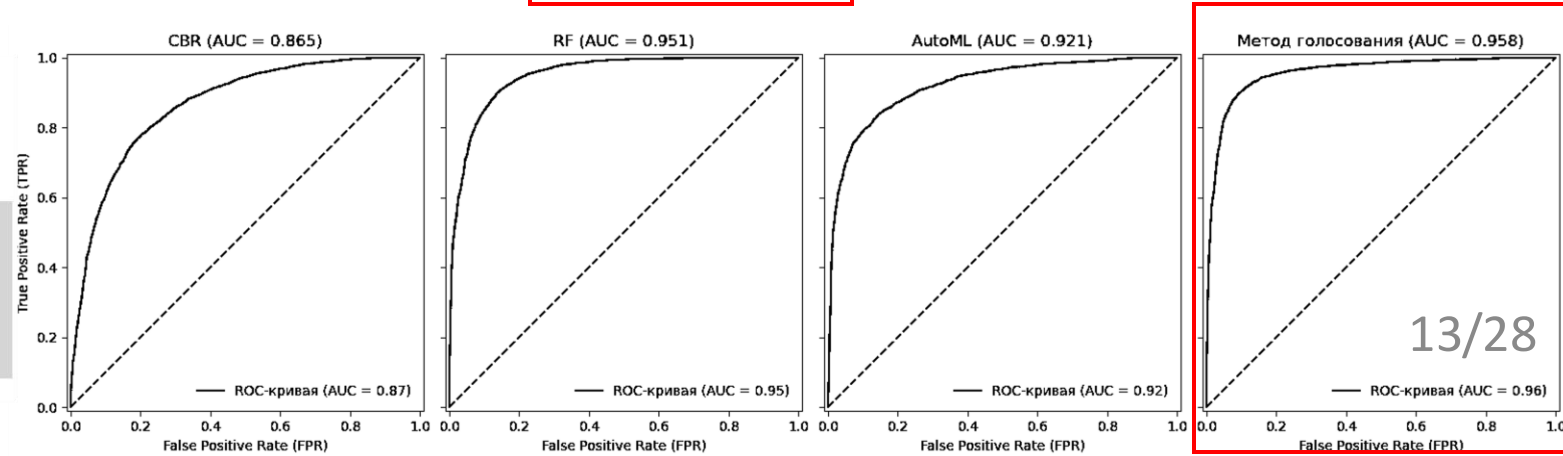


Сравнение моделей

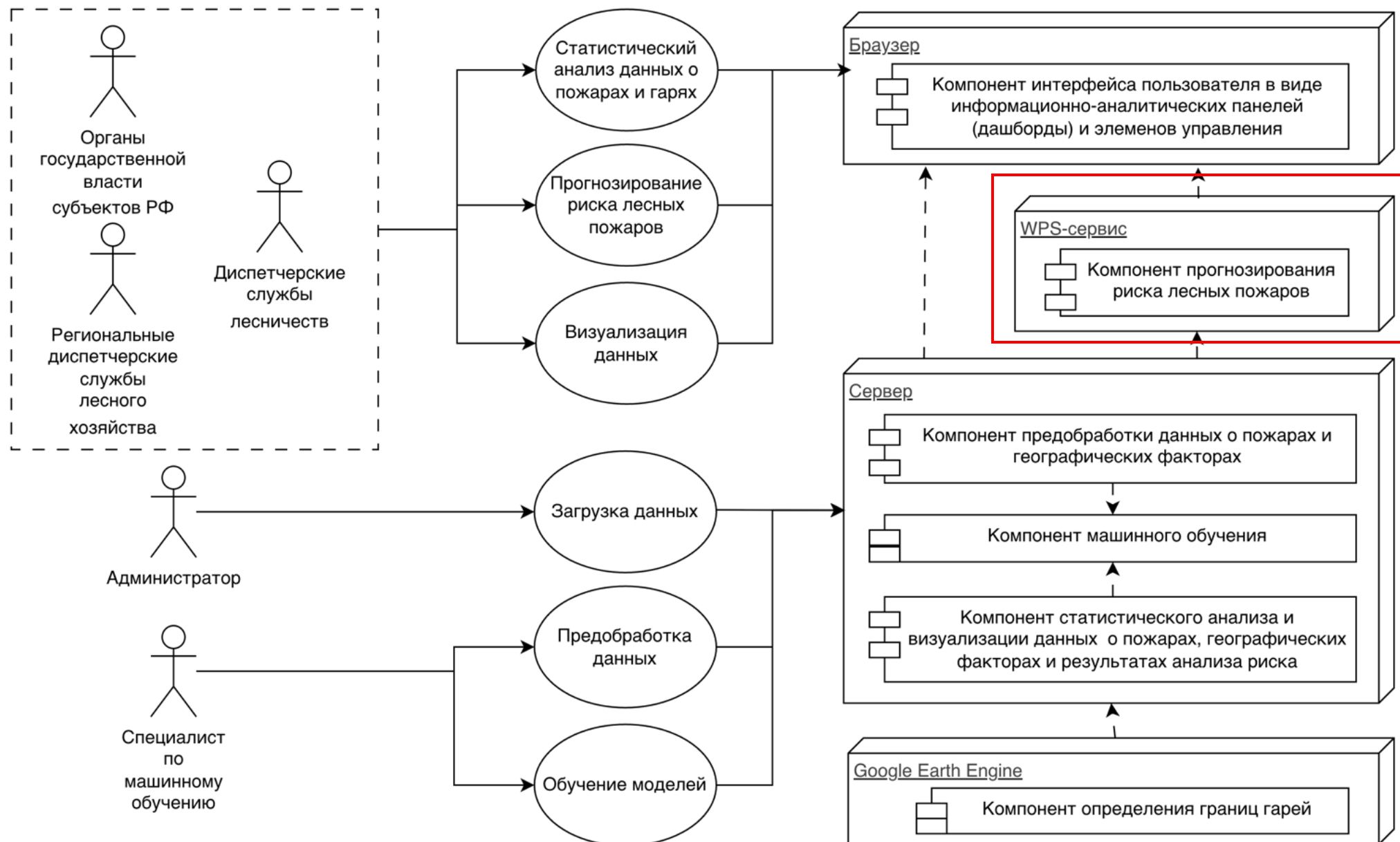


model	Метод голосования	AutoML	CBR	RF	Лучший результат
year					
2008	0.950	0.837	0.846	0.949	0.950
2016	0.849	0.861	0.767	0.826	0.861
2024	0.850	0.817	0.737	0.831	0.850
Итоговое среднее	0.894	0.843	0.795	0.883	0.894

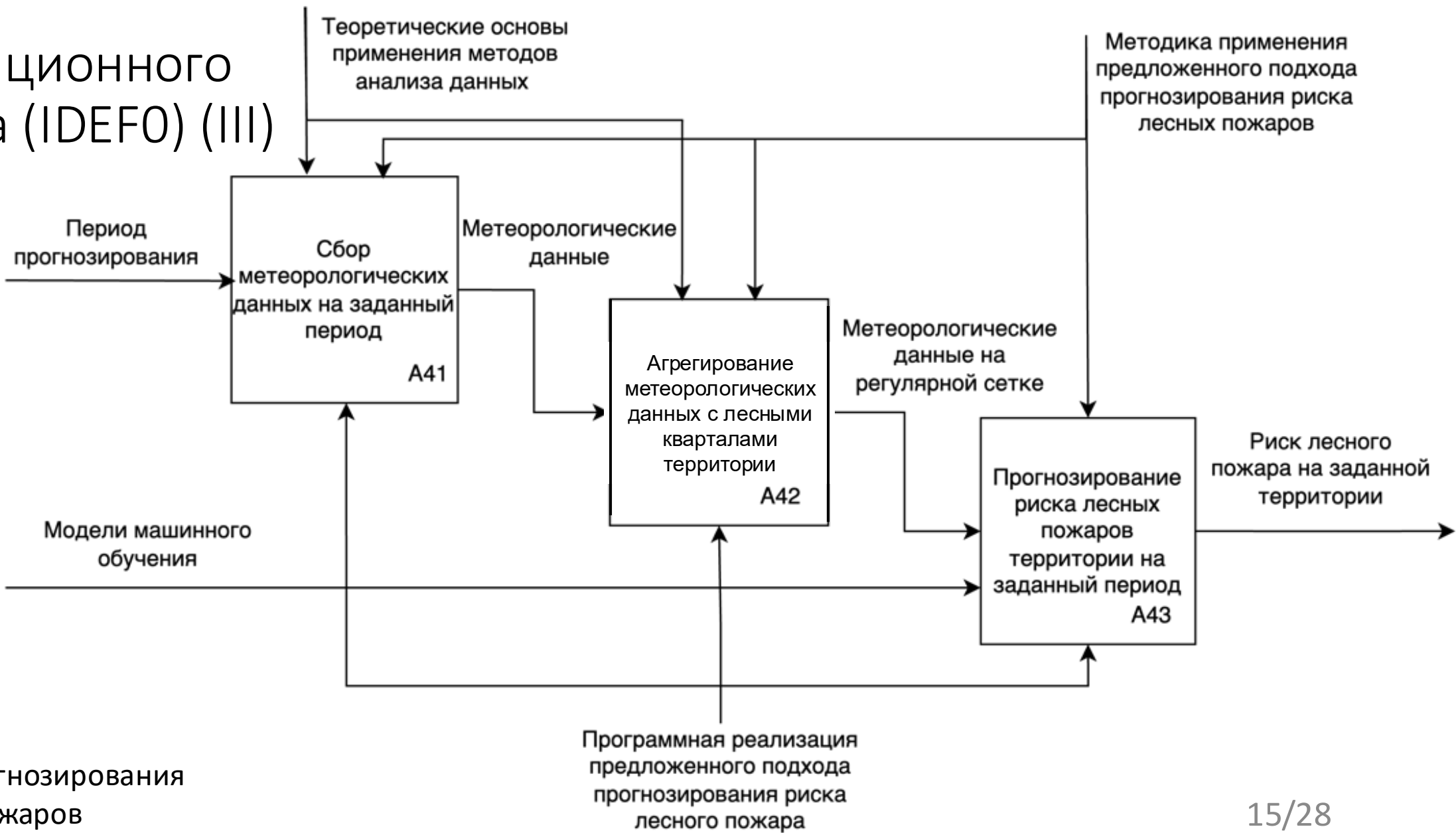
model	Метод голосования	AutoML	CBR	RF	Лучший результат
Кластер					
Кластер №1	0.834	0.835	0.717	0.803	0.835
Кластер №2	0.919	0.840	0.846	0.919	0.919
Кластер №3	0.936	0.936	0.733	0.890	0.936
Кластер №4	0.910	0.795	0.782	0.910	0.910
Итоговое среднее	0.894	0.843	0.795	0.883	0.894



Архитектура АИС



Концептуальная модель

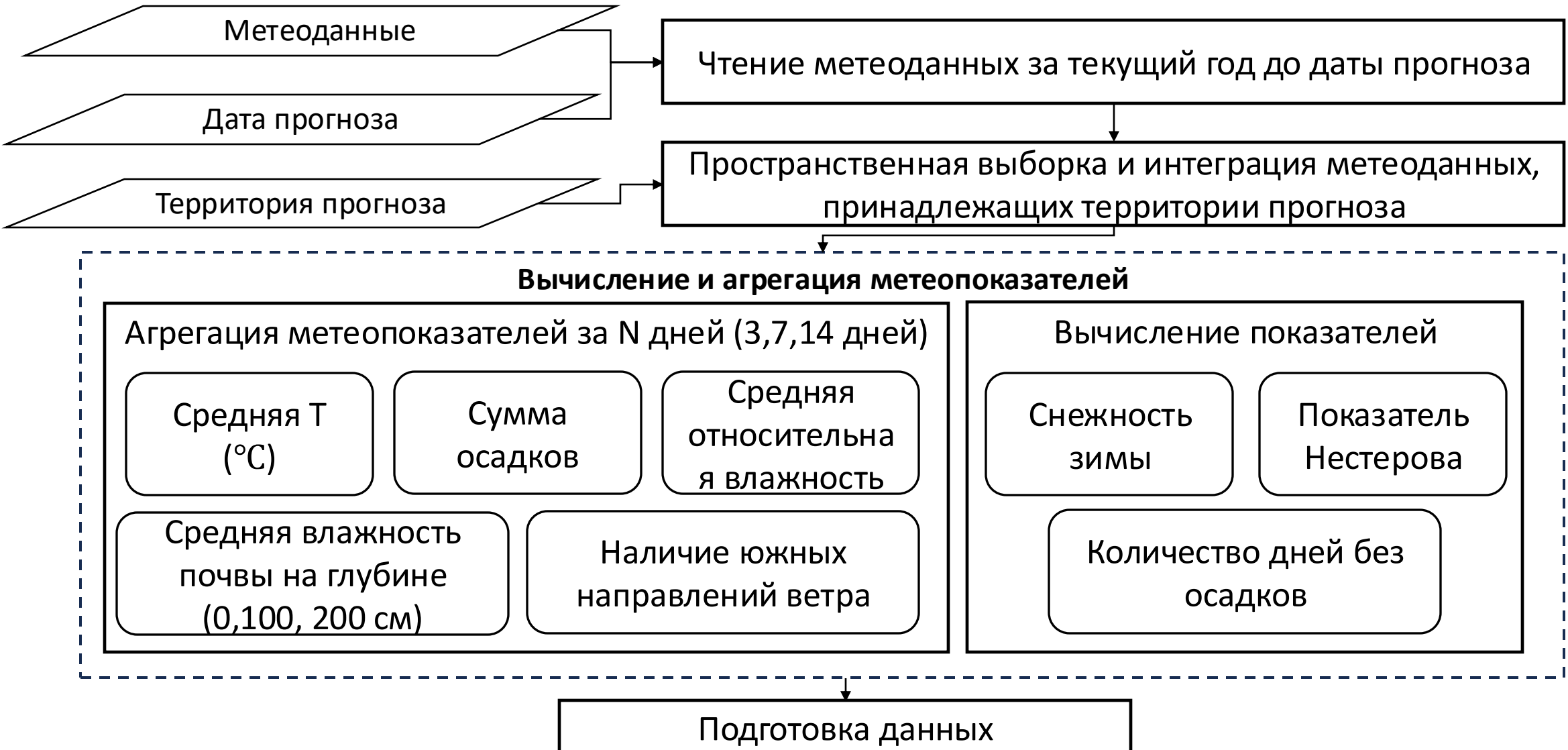


Прогнозные метеорологические данные

- Сбор и подготовка метеорологических данных для прогнозирования риска лесных пожаров на актуальный период осуществляется ежедневно автоматически с использованием продукта **GFS**, обеспечивающего прогнозирование метеорологических данных на 16 дней.
- **GFS (Global Forecast System)** – это глобальная модель прогноза погоды, разрабатываемая и поддерживаемая Национальными центрами экологического прогнозирования (NCEP), которые входят в состав Национального управления океанических и атмосферных исследований (NOAA) [1].
- Автоматизированную загрузку и обработку данных обеспечивает программный код Python с использованием модулей – requests, cfgrib, multiprocessing.

¹NCEP. Global Forecast System (GFS) [Электронный ресурс]. — Режим доступа: <https://www.nco.ncep.noaa.gov/pmb/products/gfs/> (дата обращения: 14.10.2025).

Агрегирование метеорологических данных



Функции модуля прогнозирования риска лесных пожаров



Элементы интерфейса пользователя

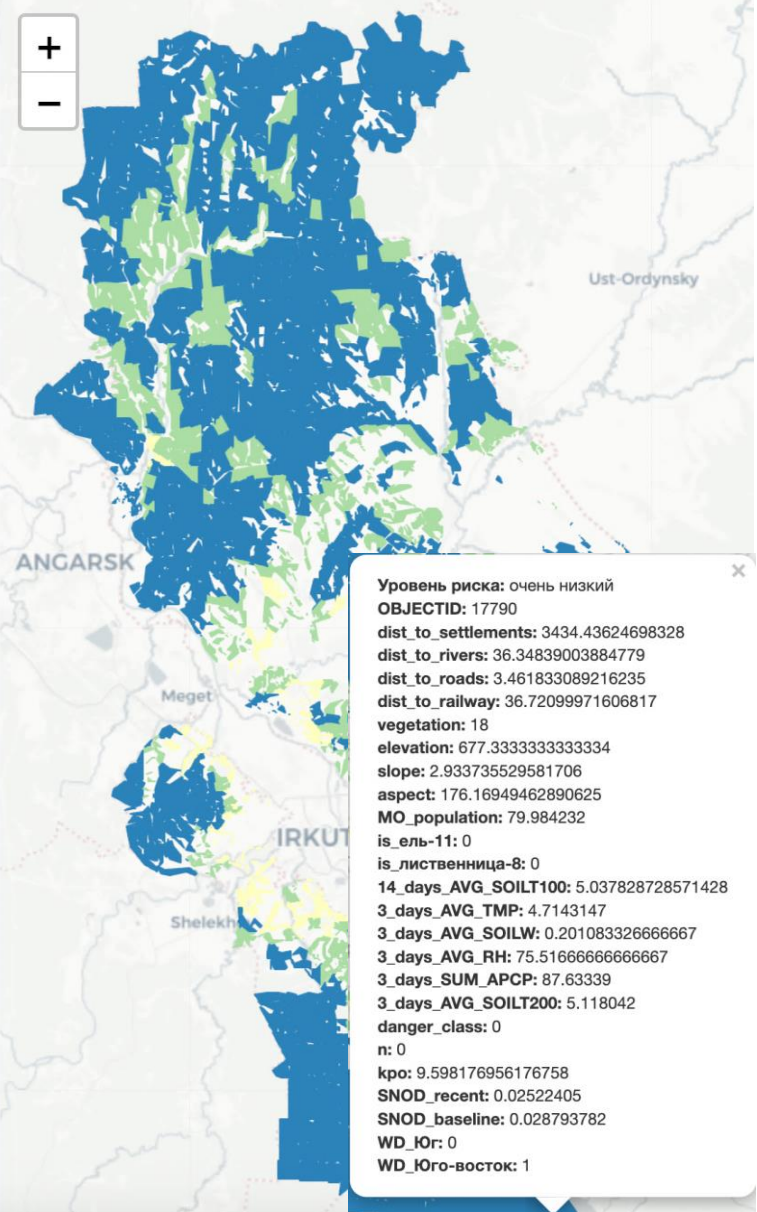
Дата прогноза: **Выбор даты**
29.10.2025

Лесничество: **Выбор территории**
Иркутское

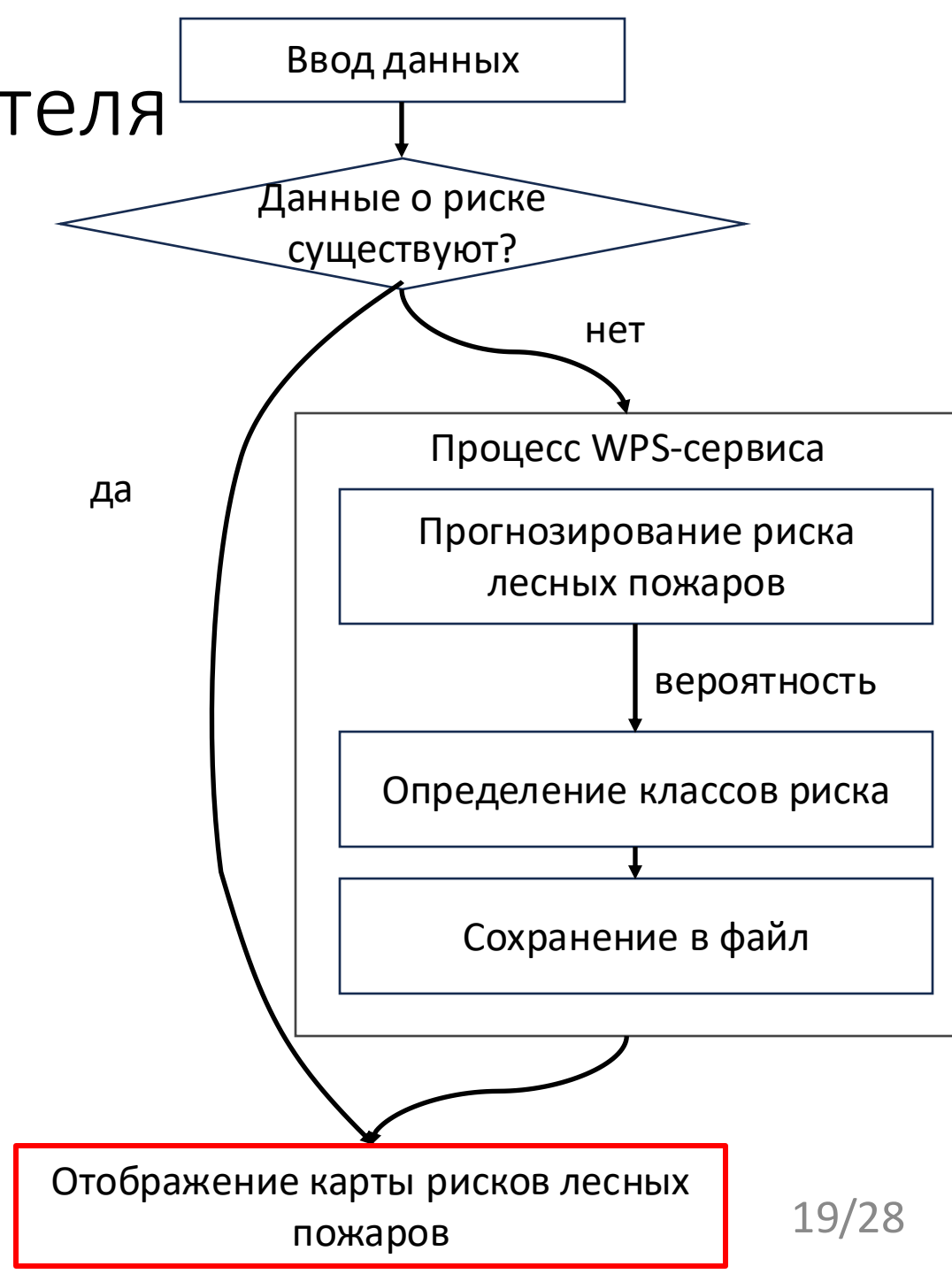
Получить прогноз

Обозначения риска:

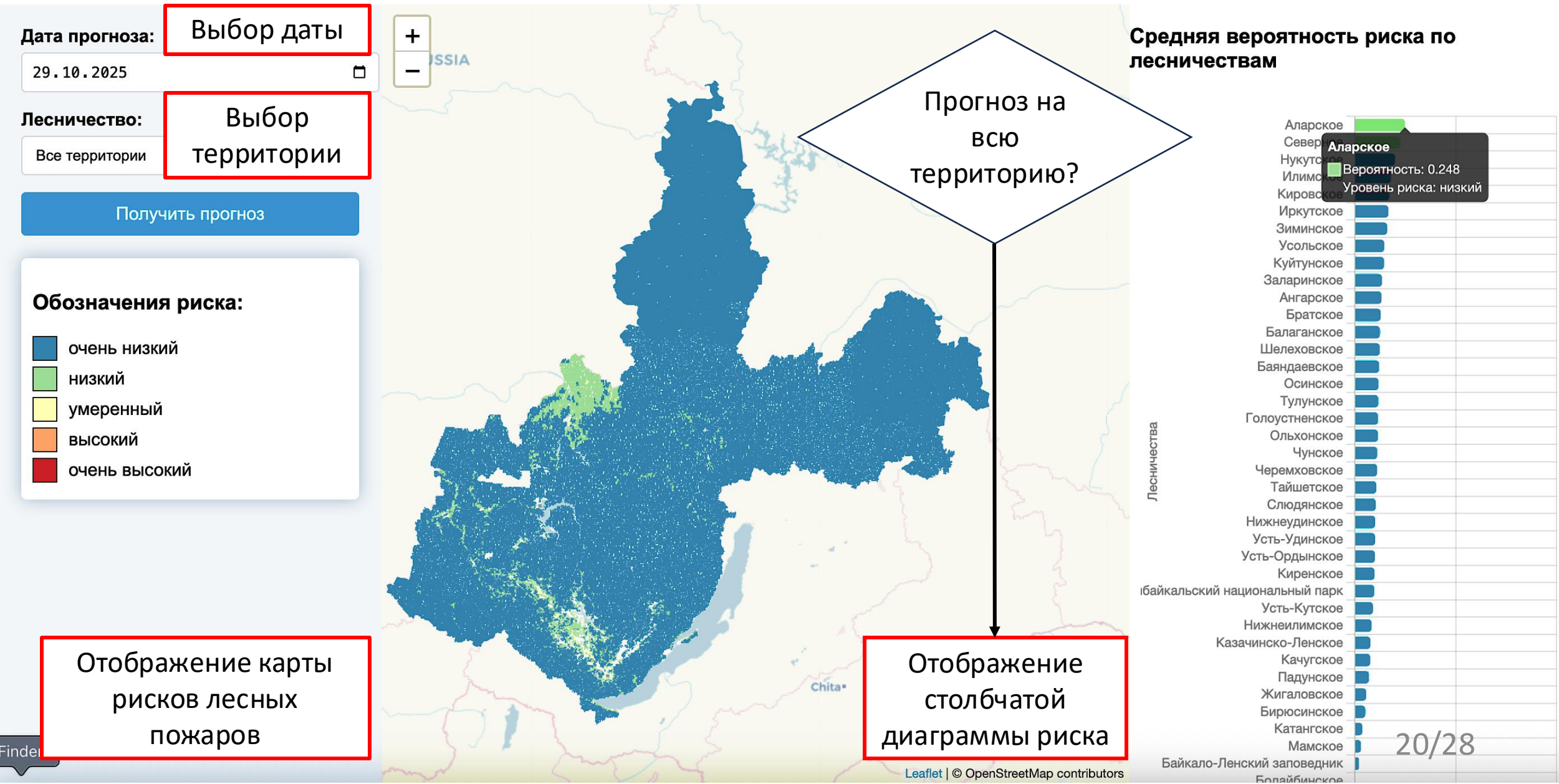
- очень низкий
- низкий
- умеренный
- высокий
- очень высокий



Уровень риска: очень низкий
OBJECTID: 17790
dist_to_settlements: 3434.43624698328
dist_to_rivers: 36.34839003884779
dist_to_roads: 3.461833089216235
dist_to_railway: 36.72099971606817
vegetation: 18
elevation: 677.3333333333334
slope: 2.933735529581706
aspect: 176.16949462890625
MO_population: 79.984232
is_ель-11: 0
is_лиственница-8: 0
14_days_AVG_SOILT100: 5.037828728571428
3_days_AVG_TMP: 4.7143147
3_days_AVG_SOILW: 0.201083326666667
3_days_AVG_RH: 75.5166666666667
3_days_SUM_APCP: 87.63339
3_days_AVG_SOILT200: 5.118042
danger_class: 0
n: 0
kpo: 9.598176956176758
SNOD_recent: 0.02522405
SNOD_baseline: 0.028793782
WD Юг: 0
WD Юго-восток: 1

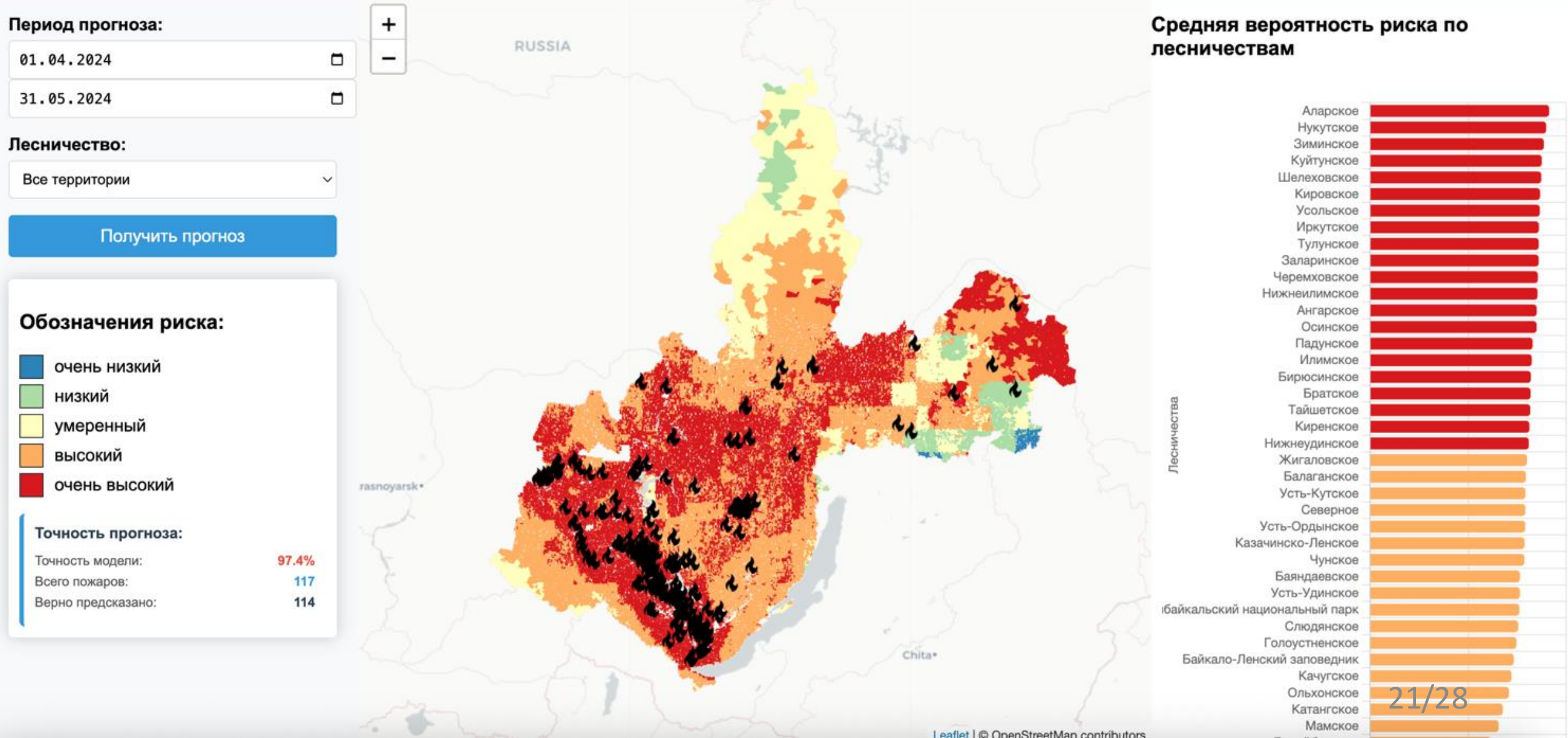


Элементы интерфейса пользователя



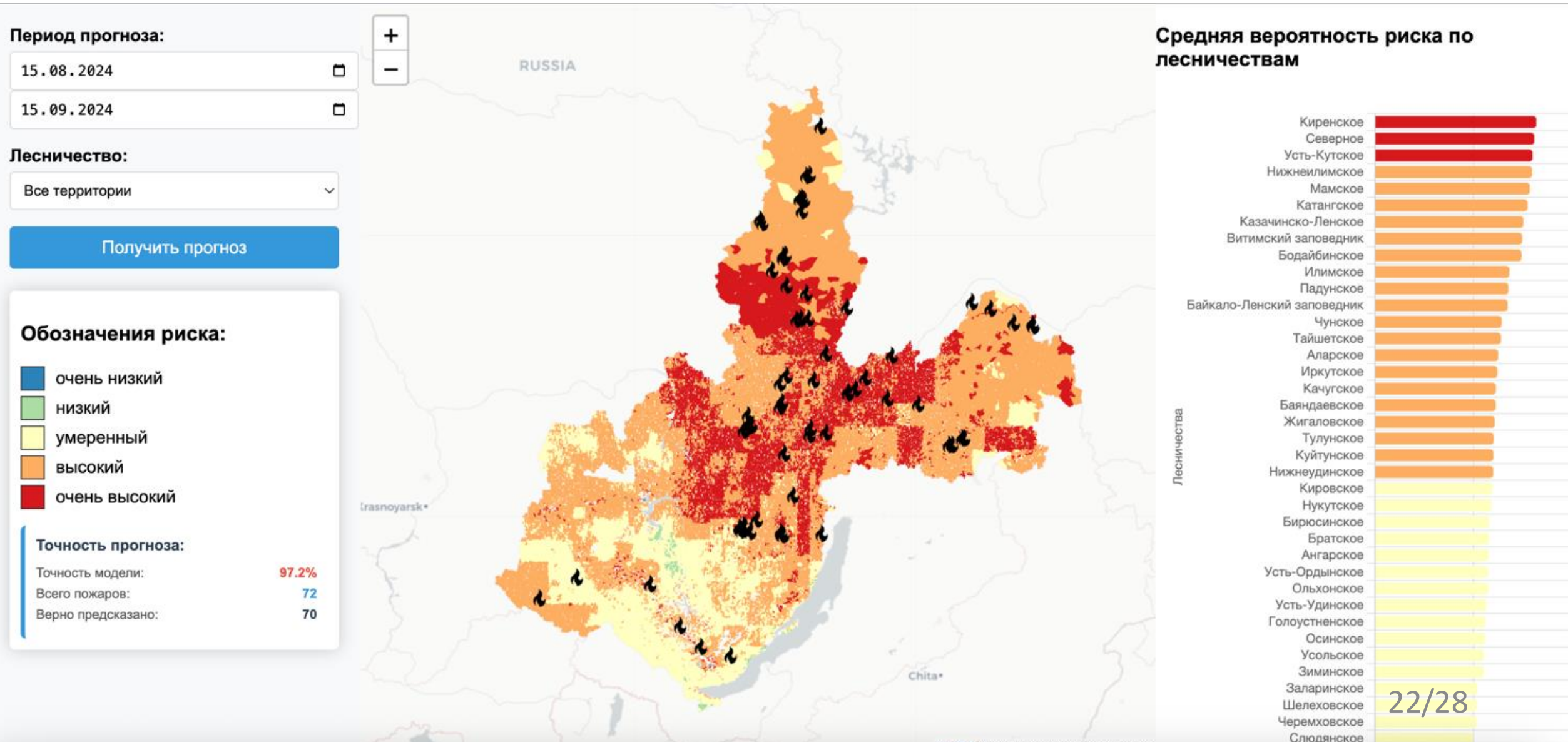
Апробация

Карта рисков лесных пожаров на территории лесных кварталов Иркутской области на весенний период с 1 апреля по 31 мая 2024 г.



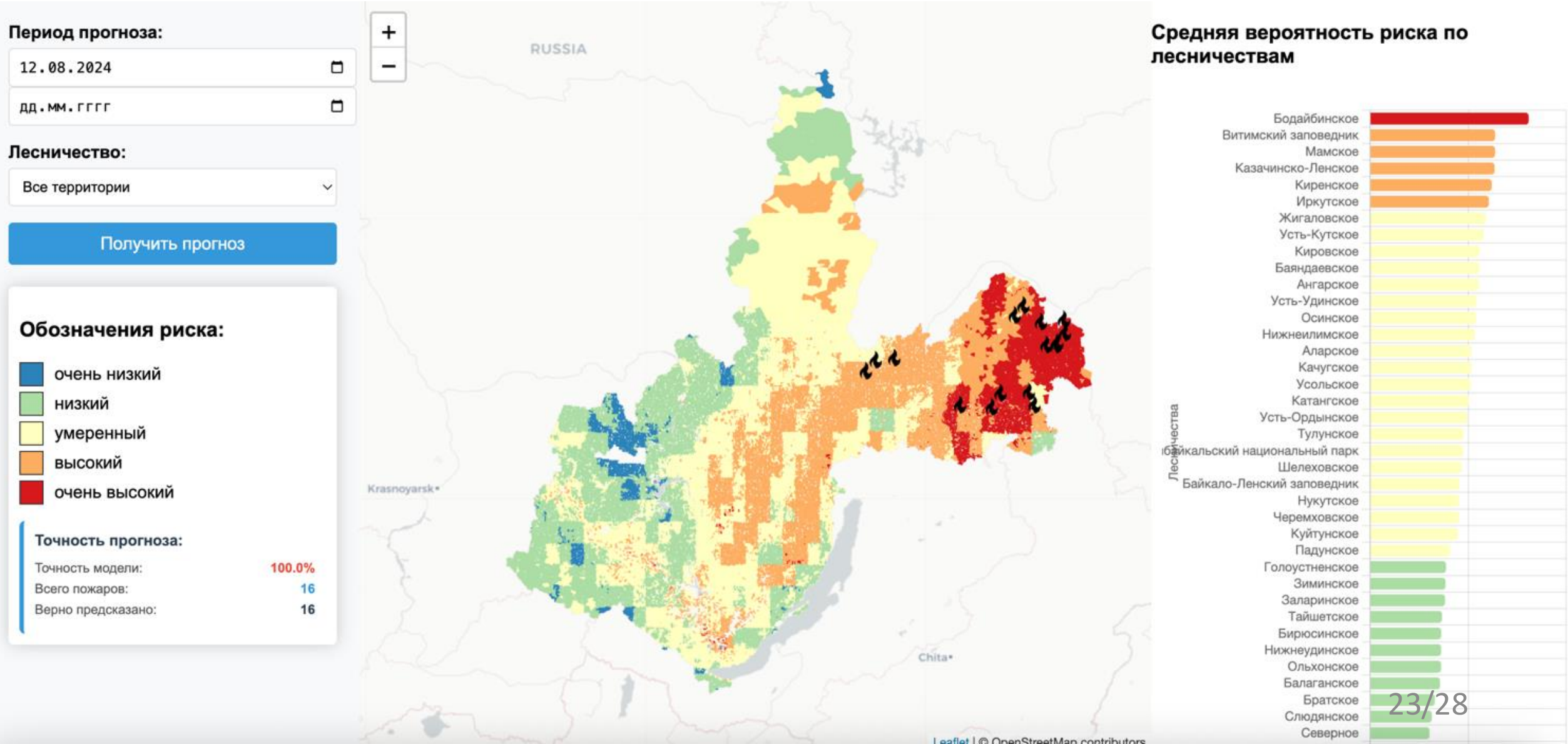
Апробация

Карта рисков лесных пожаров на территории лесных кварталов Иркутской области на период с 15 августа по 15 сентября 2024 г.



Апробация

Карта рисков лесных пожаров на территории лесных кварталов Иркутской области на 12 августа 2024 г.



Сравнение точности гибридной модели с аналогами (I)

Описание модели	Набор свойств	Метрики производительности	Эксперимент*
Модель-аналог RF для территории Якутии¹ (11 факторов / свойств)	Солнечная радиация, количество осадков, средняя и максимальная температура воздуха, тип растительного покрова, высота, крутизна и ориентация склона, дистанции до дорог, населенных пунктов и рек	F1-score: 0.8185 Accuracy: 0.8275	Recall: 0.625
Модель-аналог RF для территории Иркутской области (13 базовых факторов/ свойств)	Тип растительного покрова, высота, крутизна и ориентация склона, дистанции до дорог, населенных пунктов и рек, показатель Нестерова, количество дней без осадков, максимальная температура, количество осадков, скорость ветра	F1-score: 0.8261 Accuracy: 0.8302	Recall: 0.75

^[1] Janiec, P. A Comparison of Two Machine Learning Classification Methods for Remote Sensing Predictive Modeling of the Forest Fire in the North-Eastern Siberia / P. Janiec, S. Gadal // Remote Sens. — 2020. — Vol. 12. — P. 4157.

* Точность прогноза в результате апробации на 12.08.2024 г.

Сравнение точности гибридной модели с аналогами (II)

Описание модели	Набор свойств	Метрики производительности	Эксперимент
Модель-аналог RF для территории Иркутской области (24 значимых географических фактора/свойства)	Тип растительного покрова, наличие породы ели и лиственницы, высота, крутизна и ориентация склона, дистанции до автомобильных и железных дорог, населенных пунктов и рек, плотность населения, показатель Нестерова, класс пожарной опасности, количество дней без осадков, средняя температура, количество осадков, влажность почвы на поверхности, температура почвы на глубине 100 и 200 м, наличие южных направлений ветра, относительная снежность зимы и средняя снежность за 5 лет.	F1-score: 0.8565 Accuracy: 0.8605	Recall: 0.8125
Модель-аналог RF для территории Иркутской области с локальной балансировкой данных (24 значимых географических фактора/свойства)		F1-score: 0.883 Accuracy: 0.881	Recall: 0.875
Модель HybridML для территории Иркутской области с локальной балансировкой данных (24 значимых географических фактора/свойства)		F1-score: 0.894 Accuracy: 0.892	Recall: 1

Результаты

Разработана **гибридная модель прогнозирования риска лесных пожаров**, основанная на данных ДЗЗ и агрегации методов машинного обучения (Random Forest, Case-based reasoning, AutoML) с учетом сформированной прецедентной модели на основе наиболее значимых географических факторов территории Иркутской области, результатами верификации данных о площадях пожаров, обеспечивающая высокую оценку точности (F1-score 0.89) прогнозирования.

- **Автоматизированы процессы** сбора прогнозных метеорологических показателей, **прогнозирования риска** лесных пожаров.
- **Разработан прототип пользовательского интерфейса** для получения оперативной информации о классах риска лесных пожаров на заданную дату и территорию.
- В рамках апробации автоматизированной информационной системы выполнены прогнозирование риска лесных пожаров территории Иркутской области на заданную дату и оценка его точности $P^{ML-D-Class-Fire-HybridML}$, которая составила по метрике полноты (Recall) для весеннего периода – 0.974, для месячного периода с 15 августа по 15 сентября – 0.972, для 12 августа – 1.

Основные публикации

1. **Пестова Ю.В.** Использование прецедентного подхода для прогнозирования риска лесных пожаров / Дородных Н.О., Николайчук О.А., Пестова Ю.В., Юрин А.Ю. // Вычислительные технологии. – 2022. – Т. 27, № 5. – С. 43-54. **(БАК, УБС-4, Scopus Q4, РИНЦ).**
2. **Pestova J.** Forest Fire Risk Forecasting with the Aid of Case-Based Reasoning / Dorodnykh N., Nikolaychuk O., Pestova J., Yurin A. // Applied Sciences. – 2022. – №12(17):8761. **(УБС-2, Scopus Q1, WOS, РИНЦ).**
3. **Pestova Yu.V.** Creation of prototypes of case-based knowledge bases using transformations of decision tables to predict the risk of forest fires / Dorodnykh N.O., Nikolaichuk O.A., Pestova Yu.V., Yurin A.Yu. // Pattern Recognition and Image Analysis. – 2023. – Vol. 33, № 3. – P. 274–281. **(УБС-2, WOS Q3, РИНЦ).**
4. **Pestova Yu.V.** Wildfire Susceptibility Mapping in Baikal Natural Territory Using Random Forest / Nikolaychuk O.A., Pestova Yu.V., Yurin A.Yu. // Forests. – 2024. – Vol. 15, № 1. – P. 170. **(УБС-1, WOS Q1, РИНЦ).**

 Спасибо за внимание!

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Гибридная модель прогнозирования риска лесных пожаров,
основанная на данных ДЗЗ и агрегации методов машинного обучения
(Random Forest, Case-based reasoning, AutoML) с учетом сформированной
прецедентной модели на основе наиболее значимых географических факторов
территории Иркутской области, результатами верификации данных о площадях
пожаров, обеспечивающая высокую оценку точности (F1-score 0.89)
прогнозирования.

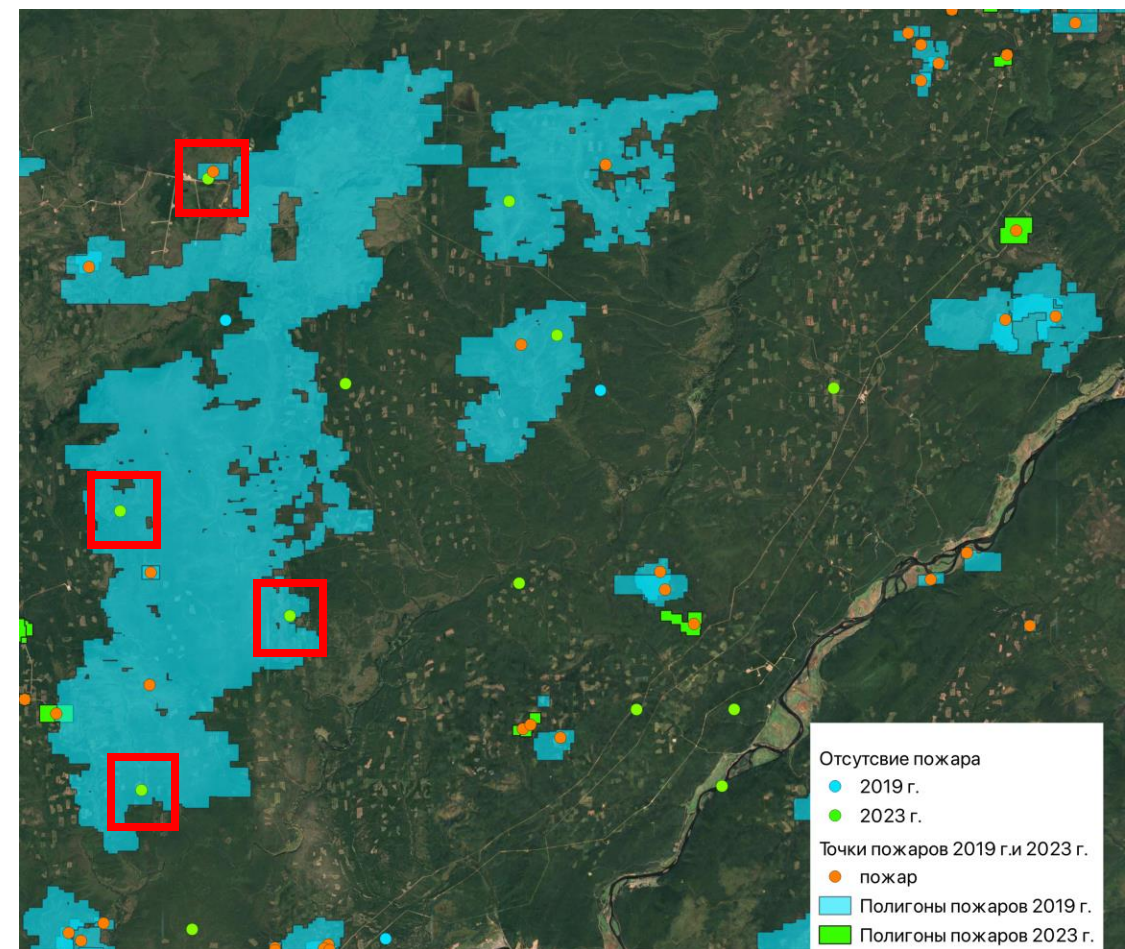
Генерация данных, для формирования класса «отсутствие пожара»

Для формирования класса «отсутствие пожара» сгенерированы случайные точки, случайность определена во времени и пространстве.

Сгенерированные координаты событий расположены на территории Иркутской области вне уточненных площадей возгорания пожаров, населенных пунктов и техногенных объектов.

Количество сгенерированных точек соответствует ежегодному количеству пожаров. Каждый класс содержит **26570** образцов.

Даты событий «отсутствие пожара» определены с учетом обратного распределения частоты возникновения пожаров по месяцам.



Исторические пожары (точки «наличие пожара») и сгенерированные (точки «отсутствие пожаров»)

Обучающая и тестовая выборки

Тестовая выборка – 8746 записей (16%) за каждые 8 лет (2008, 2016 и 2024 года).

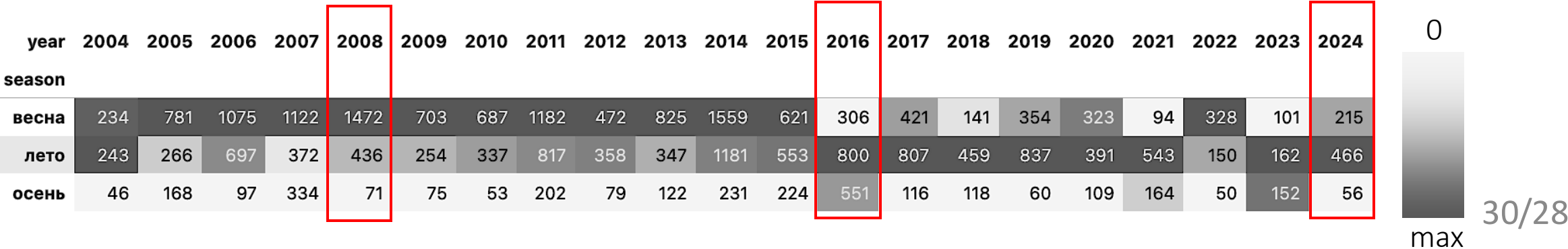
Обучающая выборка – 44394.

Классы сбалансированы и имеют соотношение идентичное исходному набору – 1:1.

Пожароопасные периоды:

- 2008 год - весенние пожары,
- 2016 – летне-осенние,
- 2024 – летние.

Количество пожаров:



Факторы, влияющие на лесные пожары

В исследованиях среди факторов влияющих на возникновение пожаров отмечают следующие группы:

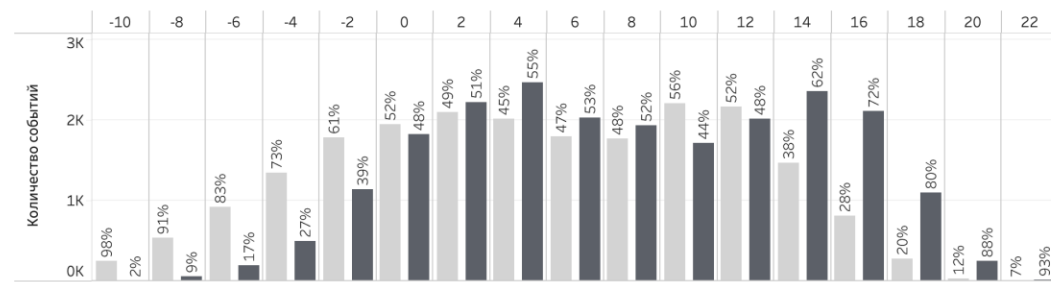
- метеорологические^{1,2},
- топографические,
- характеристики растительности и
- социальные факторы (факторы, влияющие на поведение человека).

1. Saha, S. The NCEP Climate Forecast System Reanalysis / S. Saha, S. Moorthi, H. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer et al. // Bulletin of the American Meteorological Society. — 2010. — Vol. 91, № 8. — P. 1015–1058.
2. Saha, S. The NCEP climate forecast system version 2 / S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. Chuang, M. Iredell et al. // Journal of climate. — 2014. — Vol. 27, № 6. — P.2185–2208.

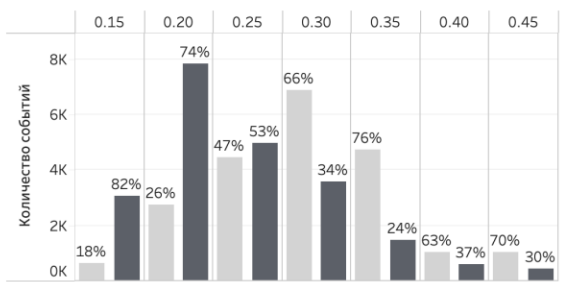
Метеорологические факторы (I)

ФАКТОРЫ	ЕД. ИЗМ.	СПОСОБ ПОЛУЧЕНИЯ:
Температура воздуха (T)	С°	Агрегирование данных RDA на регулярной сетке для временных интервалов (3, 7 и 14 дней) до начальной даты событий в исходном наборе представителей классов: «пожар» и «отсутствие пожара»
Средняя температура почвы в слоях: 0–10 см.; 10–40 см.; 40–100 см; 100–200 см. (TMPGR10, TMPGR40, TMPGR100, TMPGR200)	С°	
Средняя относительная влажность (RH)	%	
Средняя влажность почвы в слоях: 0–10 см; 10–40 см; 40–100 см; 100–200 см. (SOILW10, SOILW 40, SOILW100, SOILW200)	%	

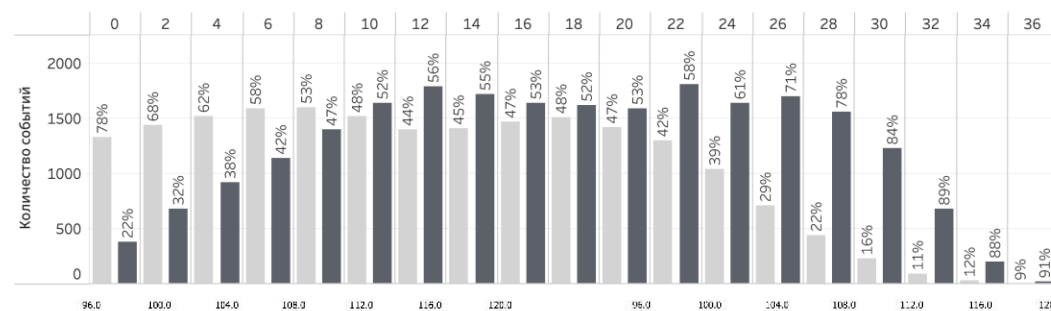
Средняя температура почвы



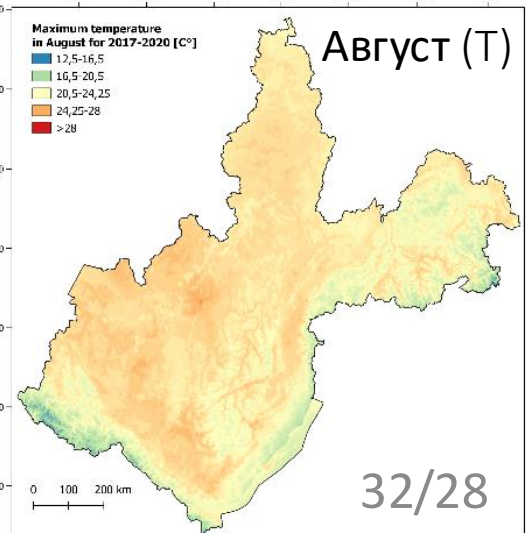
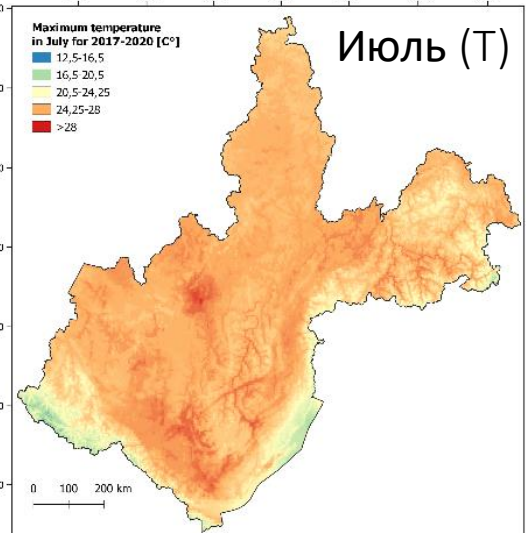
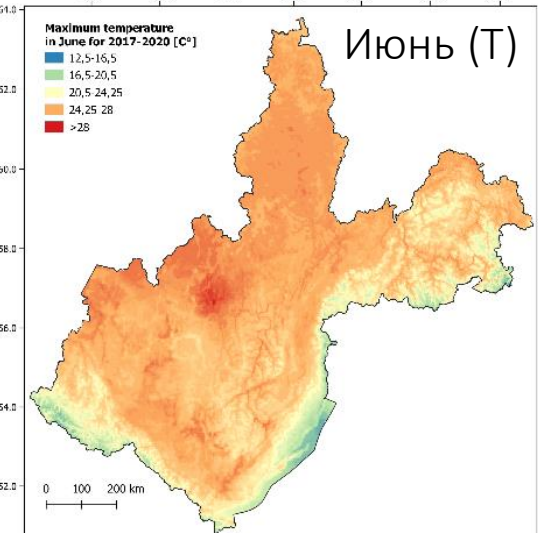
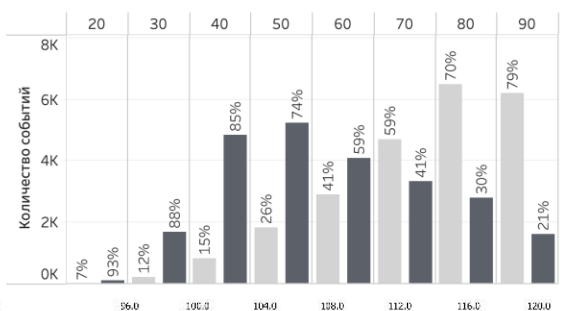
Средняя влажность почвы



Средняя температура воздуха

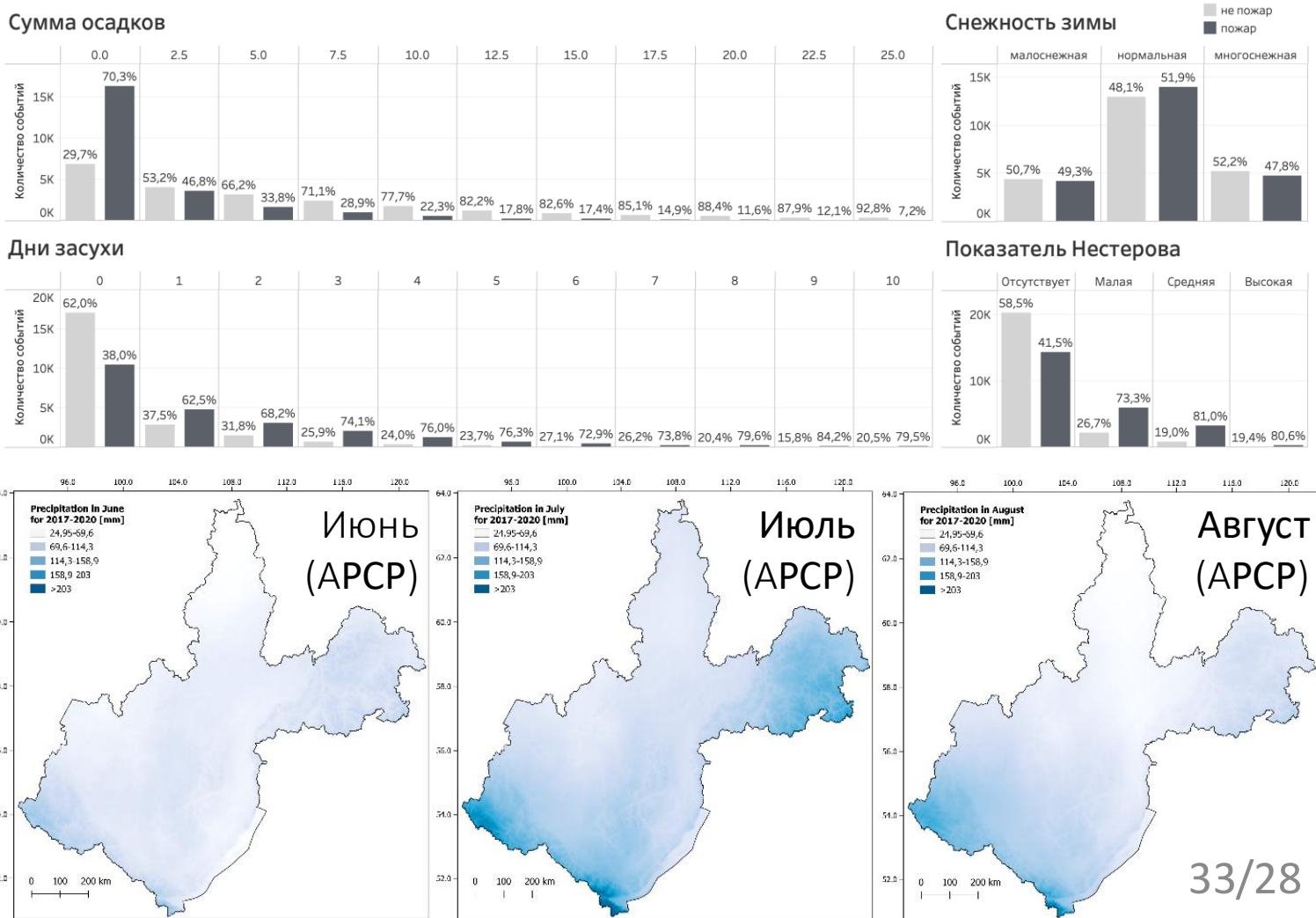


Средняя влажность воздуха



Метеорологические факторы (II)

ФАКТОРЫ	ЕД. ИЗМ.	СПОСОБ ПОЛУЧЕНИЯ:
Суммарное количество осадков (APCP)	мм.	Агрегирование данных RDA на регулярной сетке для временных интервалов (3, 7 и 14 дней) до начальной даты событий в исходном наборе представителей классов: «пожар» и «отсутствие пожара»
Показатель Нестерова (danger_class)	классы 1-5	
Снежность зимы (snowiness)	1-3 (малая, норма, много)	
Количество дней засухи: строгая, умеренная, периодическая (n)	целое число	



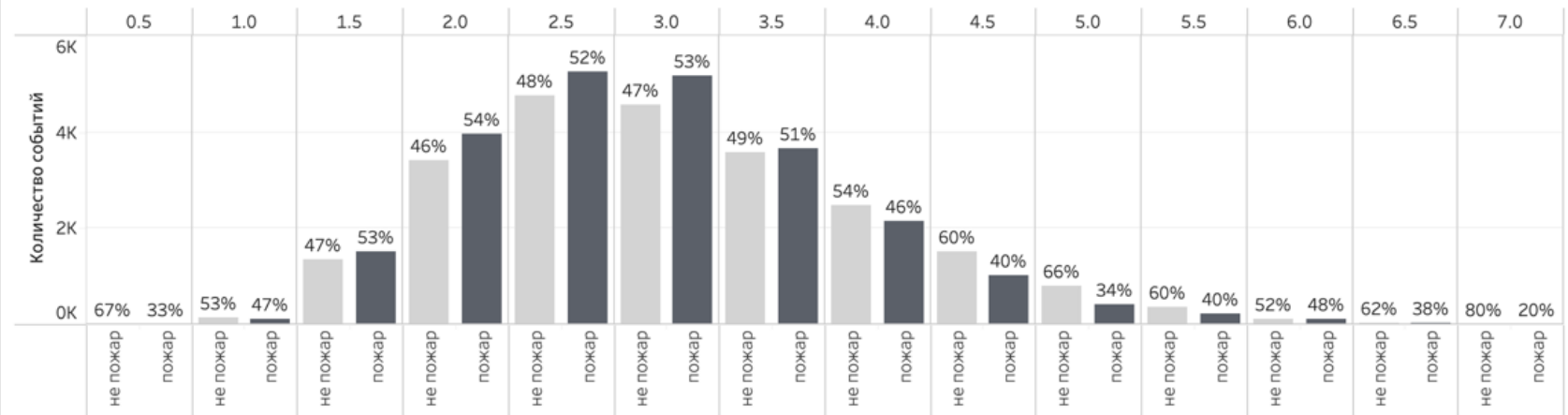
Метеорологические факторы (III)

ФАКТОРЫ	ЕД. ИЗМ.	СПОСОБ ПОЛУЧЕНИЯ:
Средняя скорость ветра (WIND_SPEED)	м/с	Агрегирование данных RDA на регулярной сетке для временных интервалов (3, 7 и 14 дней) до начальной даты событий в исходном наборе представителей классов: «пожар» и «отсутствие пожара»
Мажоритарное направление ветра (WIND_DIR)	Азимут (°)	

Распределение событий по направлению ветра



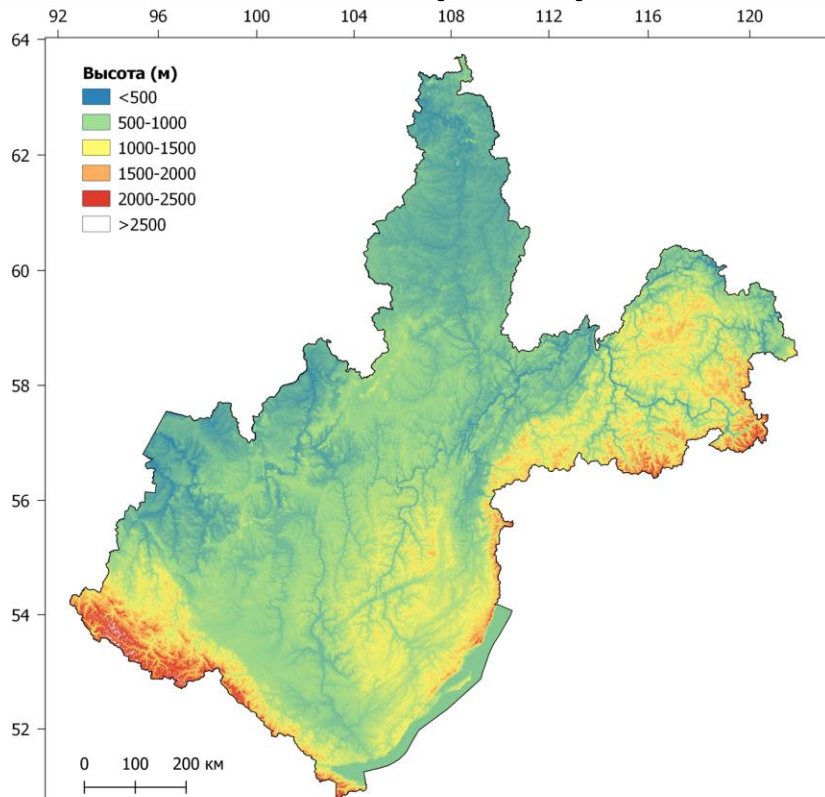
Средняя скорость ветра



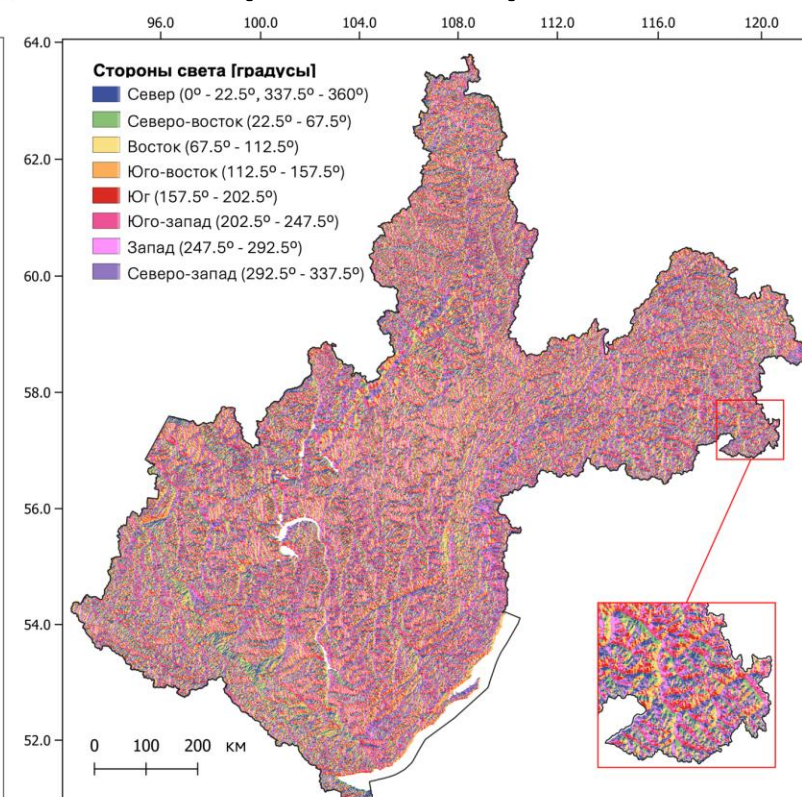
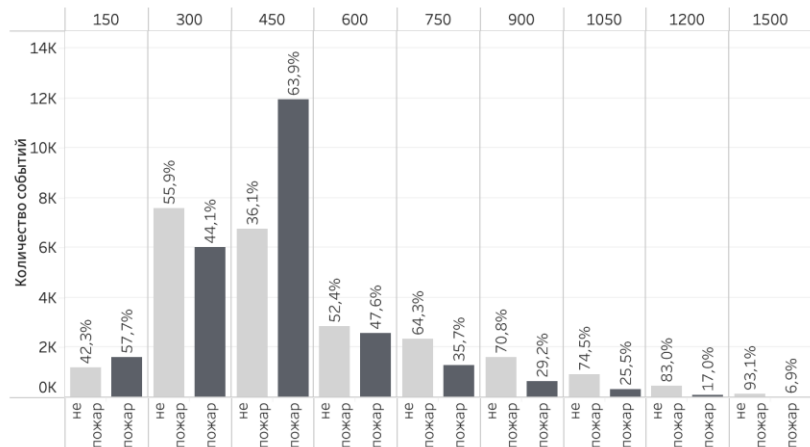
Топографические факторы (I)

ФАКТОРЫ	ОПИСАНИЕ, ЕДИНИЦЫ ИЗМЕРЕНИЯ, ЗНАЧЕНИЯ			СПОСОБ ПОЛУЧЕНИЯ
Высота над уровнем моря	метры			SRTM (https://worldclim.org/)
Аспект (ориентация наклона)	градусы (0-360)	Север	337.5-22.5	Анализ цифровой модели рельефа в QGIS
		Северо-восток	22.5-67.5	
		Восток	67.5-112.5	
		Юго-восток	112.5-157.5	
		Юг	157.5-202.5	
		Юго-запад	202.5-247.5	
		Запад	247.5-292.5	
		Северо-запад	292.5-337.5	
Уклон (крутизна)	градусы			Анализ цифровой модели рельефа в QGIS

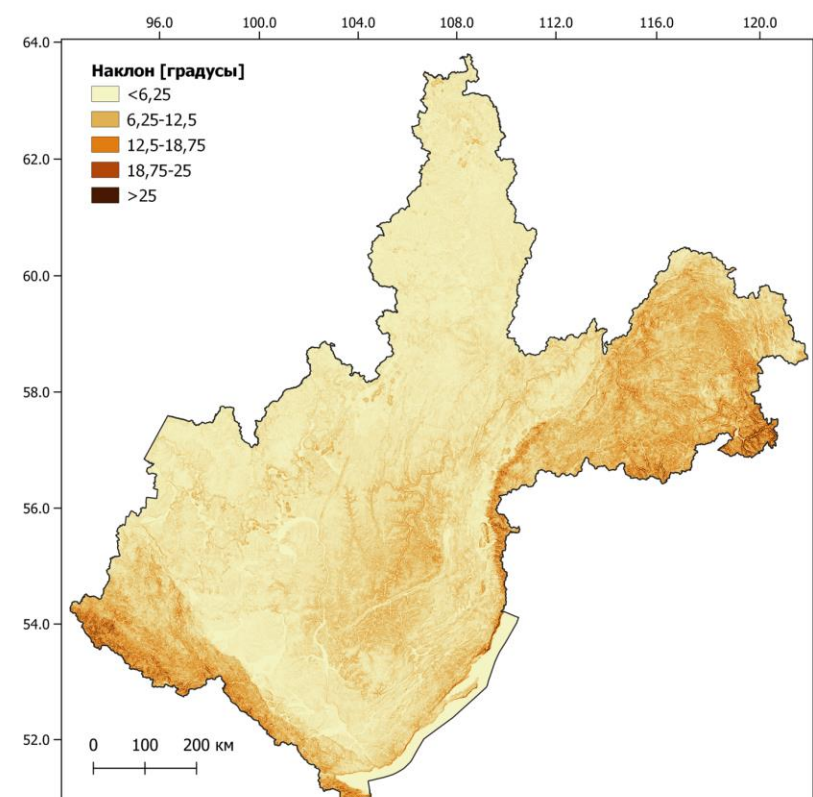
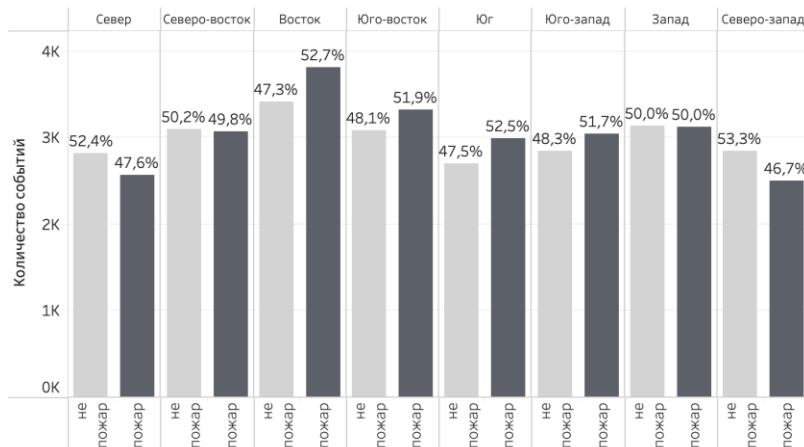
Топографические факторы (II)



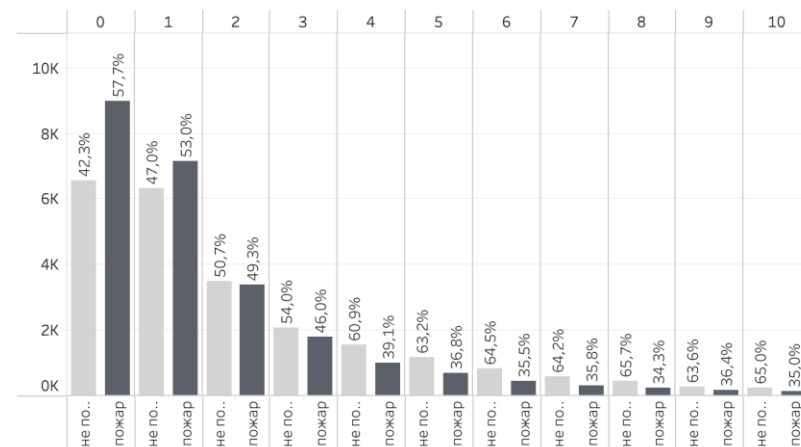
Высота



Ориентация наклона



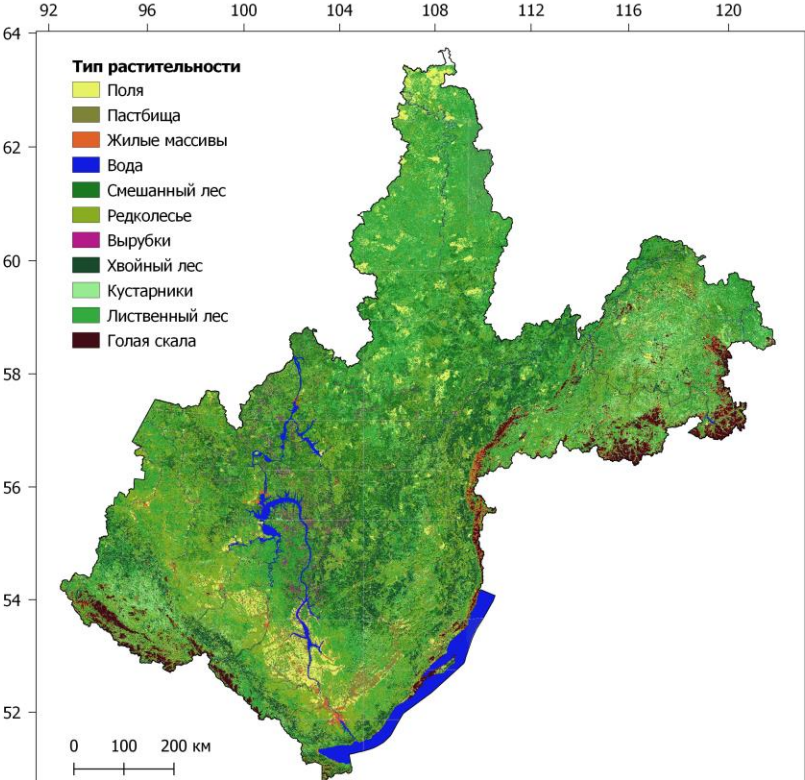
Наклон



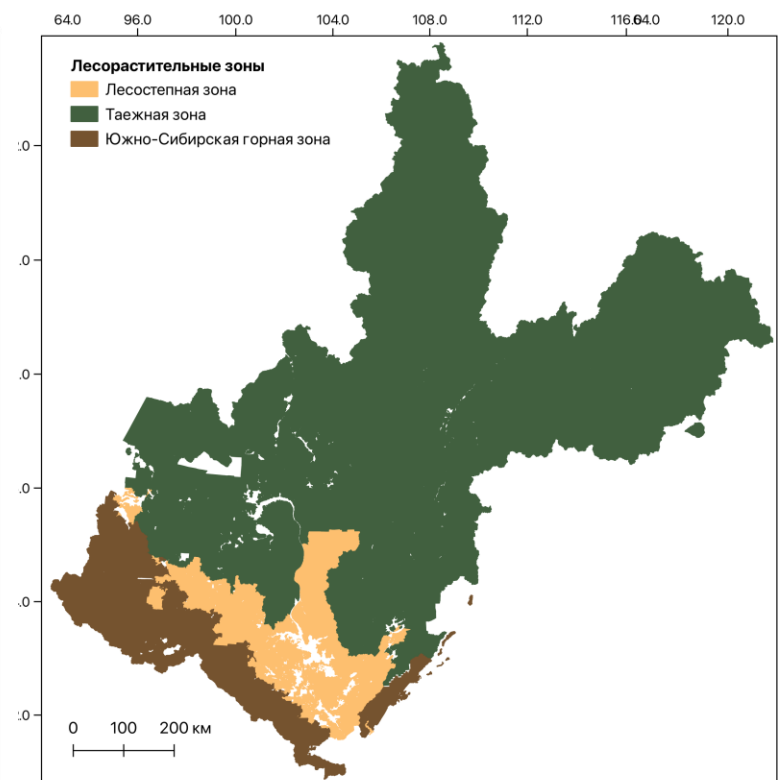
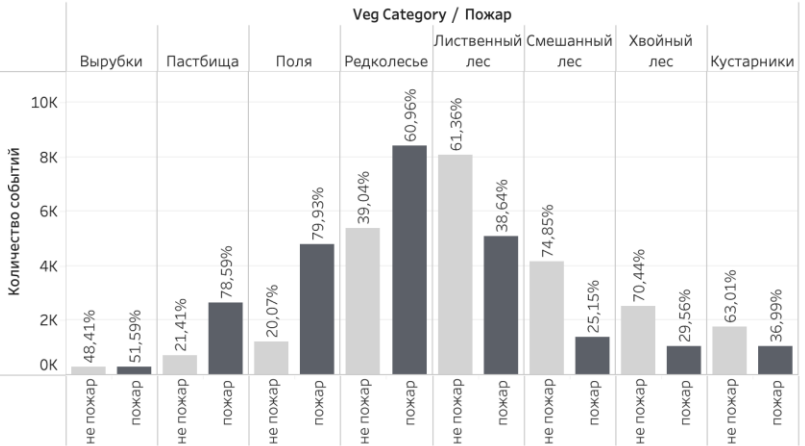
Факторы, описывающие тип поверхности (I)

ФАКТОРЫ	ОПИСАНИЕ, ВОЗМОЖНЫЕ ЗНАЧЕНИЯ	СПОСОБ ПОЛУЧЕНИЯ
Зона лесосеменного районирования (forest_seed_zoning_zones)	сосна, ель, лиственница, кедр	определяется по данным лесных регламентов
Лесорастительная зона (lesn_zone)	Лесостепная, Таежная, Южно-Сибирская горная	определяется по данным лесных регламентов
Класс опасности в лесном квартале (hazard)	5 классов пожарной опасности (1-5)	определяется по данным лесных регламентов
Класс поверхности (vegetation)	11 классов поверхности: вода, облака, жилая зона, смешанный лес, хвойный лес, лиственный лес, редколесье, голая скала, вырубки, пастбище, поля	Классификация типов поверхности земли выполнена на основе сверточной нейронной сети архитектуры ResNet-50. Использовались размеченные космоснимки Иркутской области летнего периода за 2018-2020 гг. [Бычков И.В. и др.]

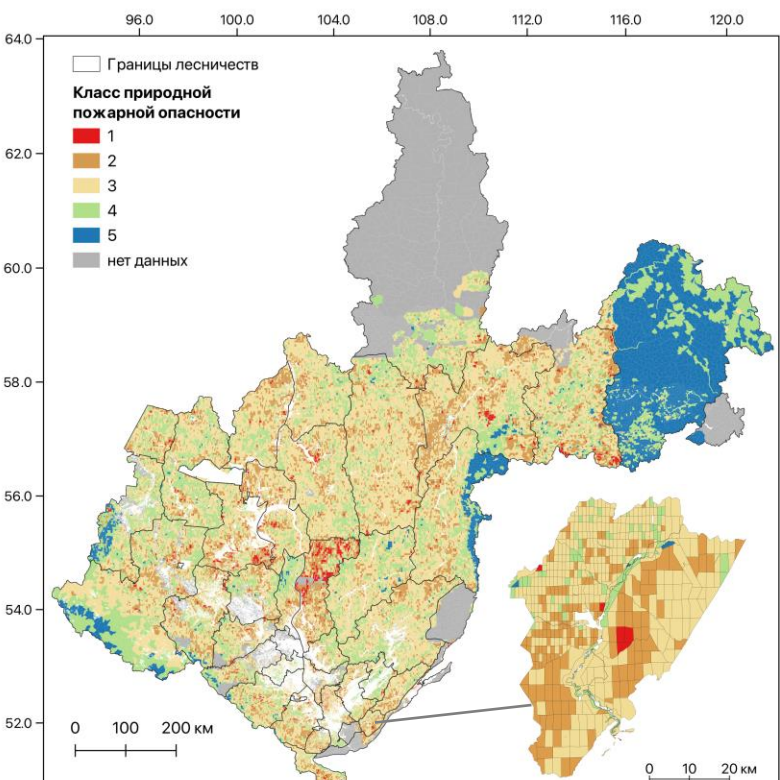
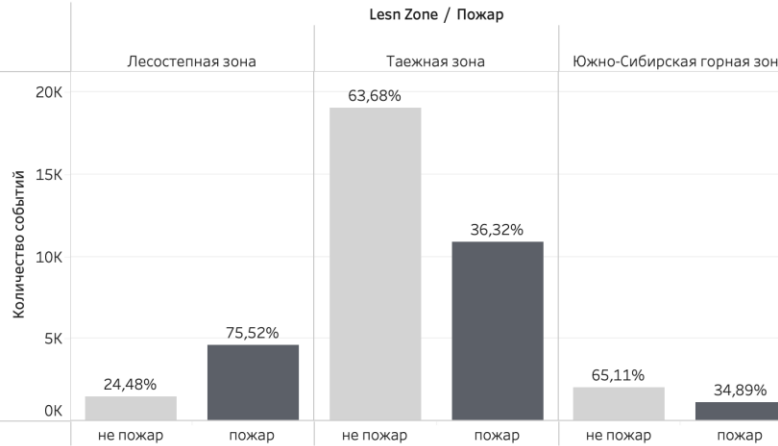
Факторы, описывающие тип поверхности (III)



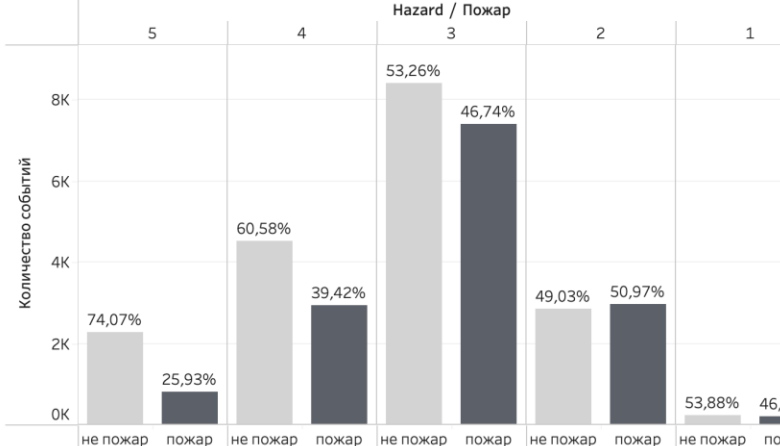
Тип растительного покрова



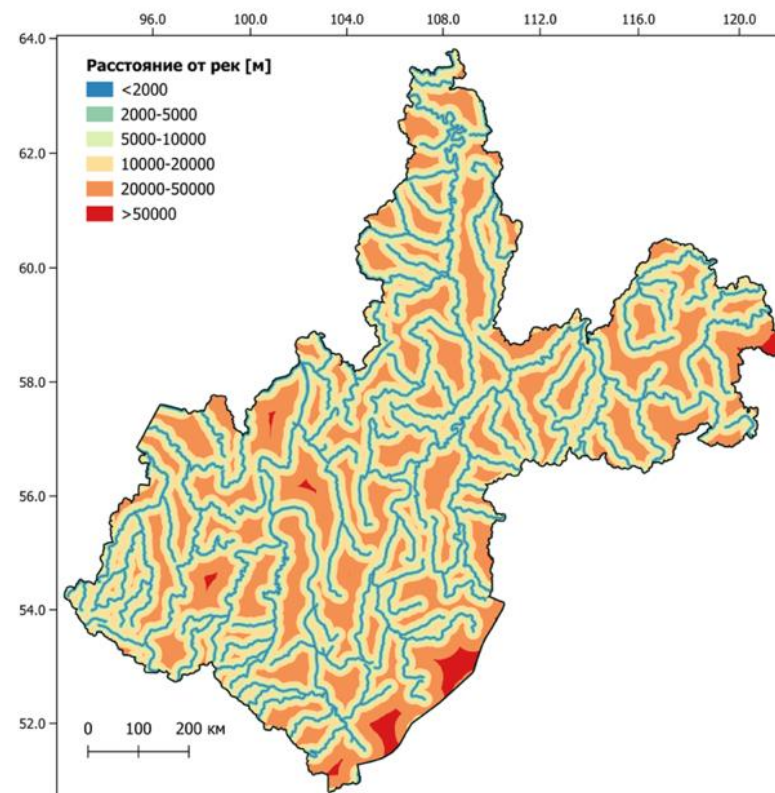
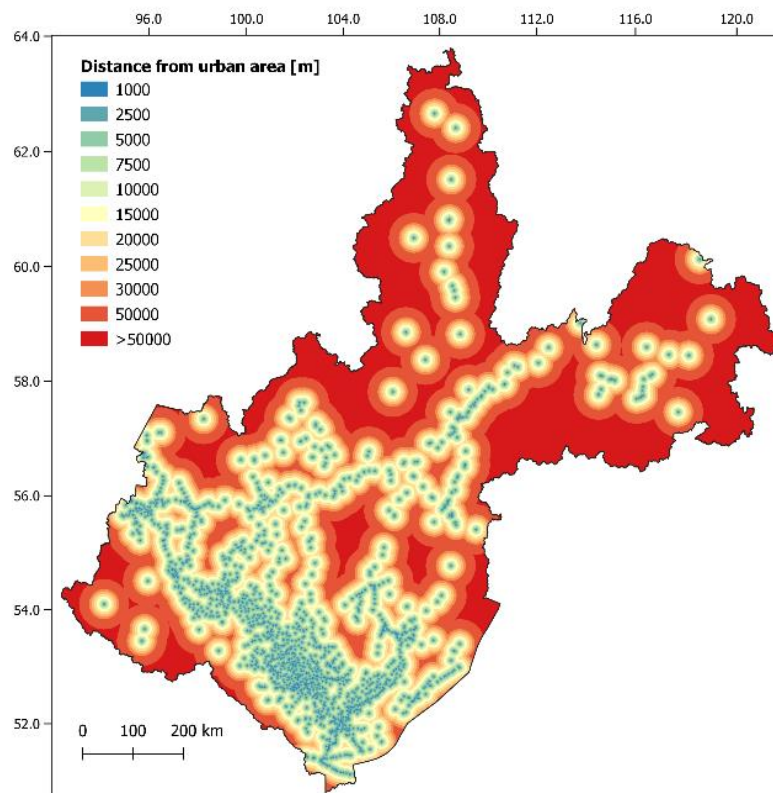
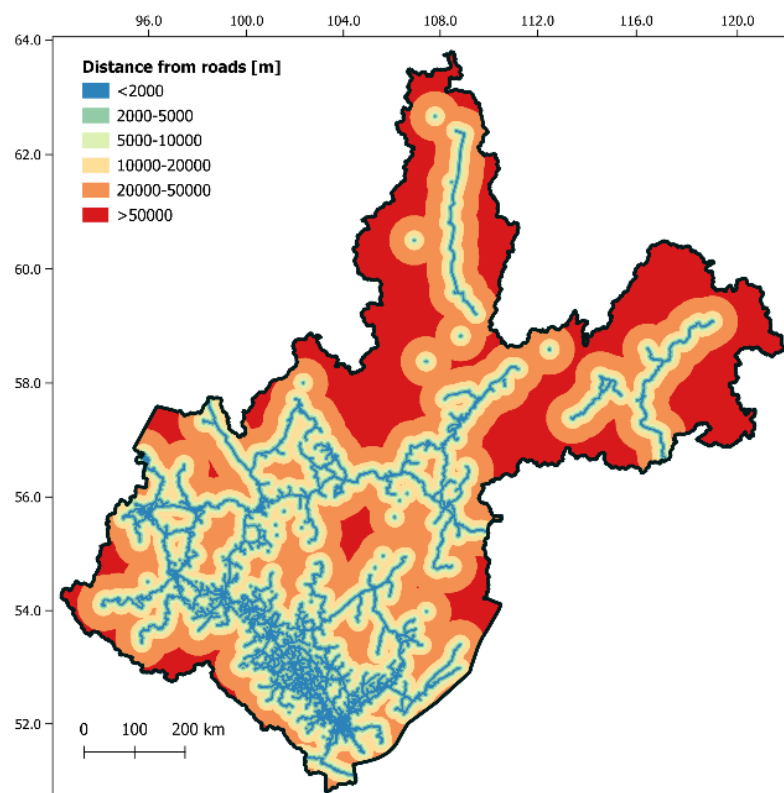
Тип лесной зоны



Класс опасности лесных кварталов



Социальные факторы (I)



ПАРАМЕТРЫ/СВОЙСТВА ПРЕЦЕДЕНТ

ОПИСАНИЕ, ЕДИНИЦЫ
ИЗМЕРЕНИЯ, ЗНАЧЕНИЯ

СПОСОБ ПОЛУЧЕНИЯ

Расстояние до ближайших объектов:
населенного пункта, автомобильной и ж/д дорог, рек

километры

определяется на основе
ГИС-данных

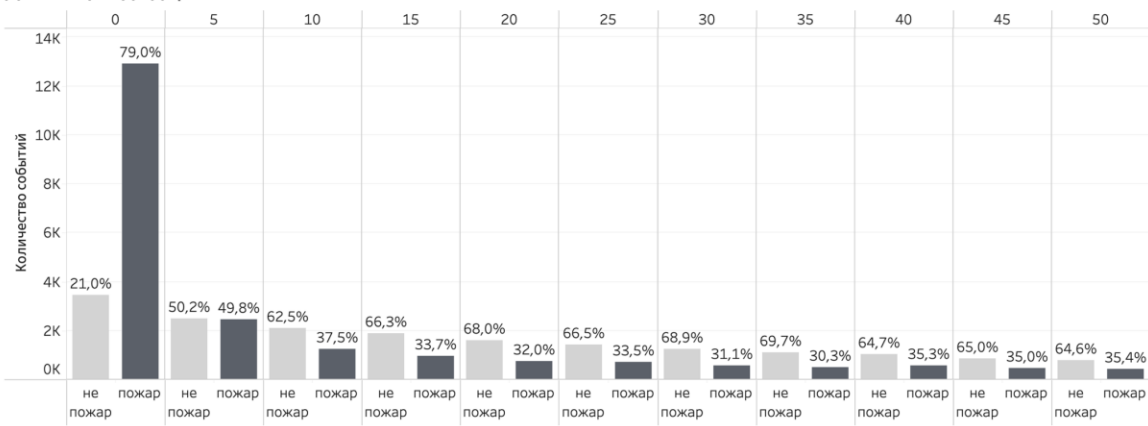
Плотность населения

чел./кв.км.

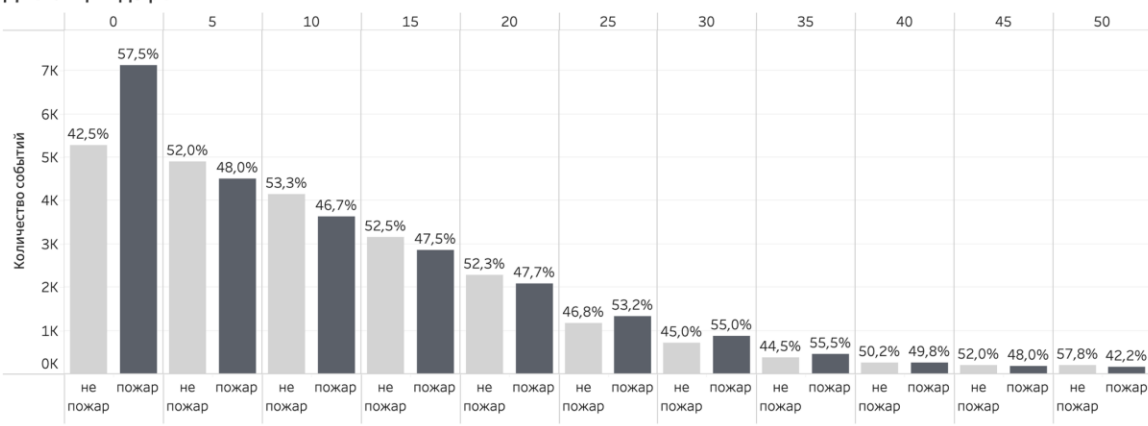
[Статистика Ирк. обл.; институт
Географии им. В.Б. Сочавы] 39/28

Социальные факторы (II)

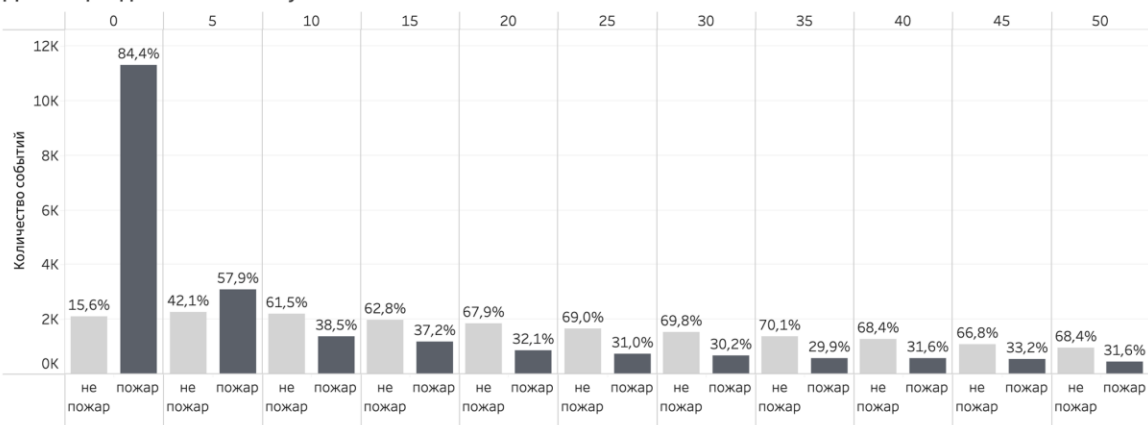
Дистанции до дорог



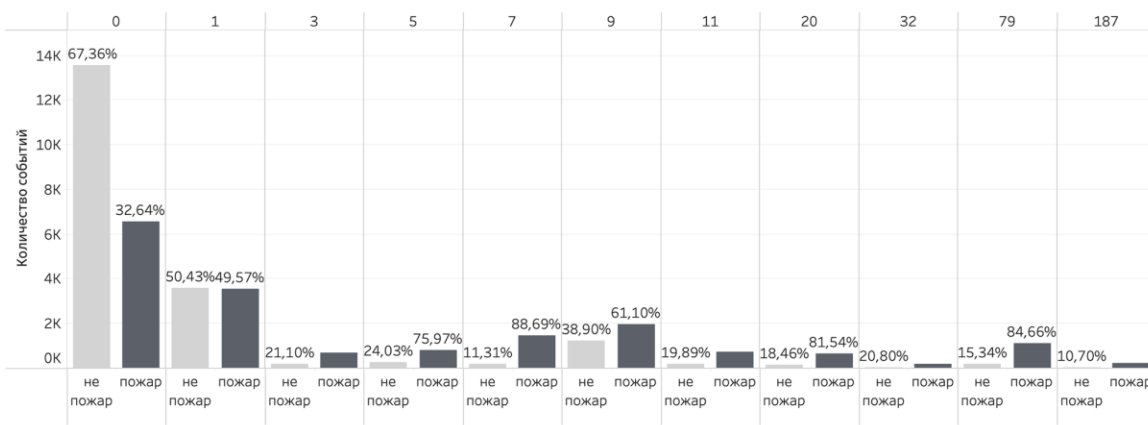
Дистанции до рек



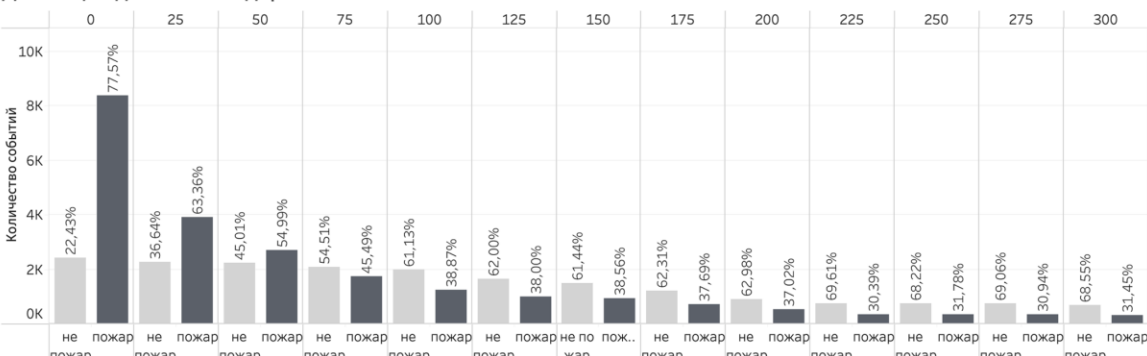
Дистанции до населенных пунктов



Плотность населения



Дистанции до железной дороги



Концептуальная модель информационного процесса (IDEF0) (I)



Концептуальная модель информационного процесса (IDEFO) (II)

