

УДК 004.852

## РАЗРАБОТКА МОДЕЛИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОГНОЗИРОВАНИЯ ПАВОДКА НА РЕКЕ ЛЕНА

### Захаров Тимур Титович

Республика Саха (Якутия), г. Якутск, Физико-технический лицей им. В. П. Ларионова, 11 класс

Научный руководитель: Романов Юрий Николаевич, г. Якутск, Физико-технический лицей им. В. П. Ларионова, учитель информатики

Наводнения – это опасное природное явления. Прогнозирование наводнений посвящено много работ. Методология нашего исследования опирается на работы Ефремова В.А., Данилов Ю.Г., Гададь С.Ж. (2019) [1], Чалов Р.С., Завадский А.С., Рулева С.Н., Кирик О.М., Прокопьев В.П., Андросов И.М., Сахаров А.И. (2016) [2], Shrakova R.N., Kusatov K.I., Mustafin S.K. (2020) [3] и модели машинного обучения.

### Этапы работы над проектом

**Первым шагом**, обобщив результаты обзора литературы и методологию исследования, мы определили, какой период реки надо брать, какие данные и что влияет на паводки.

**Второй шаг** заключался в том, чтобы собрать хорошие данные для обучения нашей модели. Для этого мы провели экспертное интервью в середине января 2021 года у и.о. начальника Якутского УГМС Мурашко Л.И. и узнали про факторы, которые влияют на заторы на реке Лена на территории городского округа “Город Якутск” и уровни контрольных точек воды во время леходода. Поработали с 15 по 20 января 2021 года в архиве ФГБУ Якутское УГМС, взяли закрытые данные для датасета. Нам показали на карте точки паводка и источники фиксирования данных (пос. Табага, г. Якутск, пос. Кангалассы).

**Третьим шагом** мы обработали данные с помощью библиотеки **Pandas** и визуализировали с помощью библиотеки **Matplotlib**. После обработки массива начали выбирать модели.

Год	Пункт	Уровень подвиек	Даты подвиек	Уровень начала ЛДХ	Дата начала лдх	Уровень наивысший ЛДХ	Дата наивысшего лдх	Продолжительность ЛДХ	Макс. Тл	Дата макс. Тл	Средняя температура	Среднее давление	
0	2008	Табага	748	733180	860	733181	877	733182	9	123	733152	7.1	1000.6
1	2009	Табага	721	733537	800	733542	823	733541	23	168	733517	7.6	991.1
2	2010	Табага	656	733910	1020	733911	1159	733912	9	156	733882	9.7	997.8
3	2011	Табага	553	734267	662	734635	935	734273	8	118	734216	9.5	997.8
4	2012	Табага	674	734639	872	735005	935	734642	10	123	734582	9.6	999.8

Рис. 1. Обработанный датасет

Далее, обработав указанный массив данных (рис. 1), мы разделили датасет на обучающую и тестовые выборки с помощью библиотеки **sklearn** и метода **train\_test\_split** с 20% тестовых данных. В итоге выбрали 3 линейные модели: Линейная Регрессия, Случайный лес и дерево решений. И получили результаты по двум метрикам (**R2** и **MAE**). Наши результаты представлены в таблице 1.

Таблица 1. Результаты моделей

	R2	MSE
Линейная регрессия	0.821	15839.86
Дерево решений	0.647	31399.79
Случайный лес	0.657	30489.96

Из таблицы 1 видно, что лучше всех показала результат линейная регрессия.

$Y = \Sigma(a + bx)$  – формула линейной регрессии

```
3 # создание модели
4 def createLSTM(units):
5     model = Sequential()
6     model.add(LSTM(units, activation='relu', input_shape=(1,13)))
7     model.add(Dropout(0.2))
8     model.add(Dense(units, activation='tanh'))
9     model.add(Dropout(0.2))
10    model.add(Dense(12))
11    model.compile(optimizer='adam', loss='mean_squared_error')
12    return model
13 # обучение модели
14 def fitLSTM(model):
15     history = model.fit_generator(train_generator, steps_per_epoch=1, epochs=500, verbose=0)
16     return history
```

Рис. 2. Создание модели LSTM

Для дальнейшего обучения на наших данных мы выбрали модель рекуррентной нейронной сети, а точнее ее разновидность **LSTM** (Long short-term memory) – особая разновидность архитектуры рекуррентных нейронных сетей, способная к обучению долговременным зависимостям (рис. 2). С помощью библиотеки **Keras**, мы сделали простую **lstm** модель, состоящую из двух слоев с функциями активации **relu** и **sigmoid** и функцией потерь – **MSE**. В отличие от других алгоритмов машинного обучения, рекуррентные нейронные сети с долгой кратковременной памятью способны автоматически выявлять признаки из временных последовательностей, обрабатывать многомерные данные, а также выводить последовательности переменной длины, благодаря чему их можно использовать для интервального прогнозирования, что упростит более точное прогнозирование (таблица 2).

Таблица 2. Результаты LSTM

	<b>R2</b>	<b>MSE</b>
LSTM	0.862	10372.2

### Серия экспериментов модели

```
1 def make_predict(model):
2     p = model.predict_generator(test_generator)
3     # print(p)
4     p = scaler.inverse_transform(p)
5     return p
6
7 def make_forecast(model, forecast_range):
8     forecast = []
9     for x, y in test_generator:
10        p = model.predict(x)
11        future_x = x
12        for step in range(forecast_range-1):
13            future_x = np.roll(future_x, -1)
14            future_x[0][-1] = p
15            p = model.predict(future_x)
16            forecast.append(p[0])
17    return scaler.inverse_transform(np.array(forecast))
```

Рис. 3. Результаты краткосрочного прогнозирования

После обучения и проверки модели мы сохранили ее и начали делать цифровой сервис (продукт) – веб приложение, где можно визуально фиксировать результаты прогноза нашей модели. Мы проверили серию экспериментов с визуализацией результатов на веб-приложении. Первый эксперимент – проверка уже известных параметров в 2018 году, чтобы протестировать гипотезу и посмотреть уровень достоверности получаемых результатов. Далее мы провели эксперименты на других датах на 2019 и на 2020 годы. Проверить достоверность полученных данных можно как отношение среднего квадратического отклонения эмпирических точек от установленной

зависимости. В результате мы получили погрешности между данными нашей модели и результатами, которые были зафиксированы Якутским УГМС. Мы сделали прогноз на 2021 и на 2022 годы. Таким образом, результаты нашего исследования показывают, что, используя большой массив данных, можно составить модель прогноза паводка на реках.

Список литературы:

1. Ефремова В.А., Данилов Ю.Г., Гадаль С.Ж. Заторные наводнения на реке Лена на территории городского округа г. Якутск // Материалы международной конференции «Холод как преимущество. Города и криолитозона: традиции, инновации, креативность». Якутск, 2019. Вып. № 2(20), С. 1-5.
2. Чалов Р.С., А.С. Завадский А.В., Рулева С.Н., Кирик О.М., В.П. Прокопьев В.П., Андросов И.М., Сахаров А.И. Морфология, деформации, временные изменения русла р. Лены и их влияние на хозяйственную инфраструктуру в районе г. Якутска // Геоморфология. 2016. № (3), С. 22-35. <https://doi.org/10.15356/0435-4281-2016-3-22-35>.
3. Shpakova R.N., Kusatov K.I., Mustafin S.K., Spatiotemporal Trends in Changes in the River Water Contents in the Sakha Republic (Yakutia) // IOP Conf. Series: Earth and Environmental Science 459 (2020) 052062 IOP Publishing doi:10.1088/1755-1315/459/5/052062.