

Предсказание рисков наводнений с помощью методов машинного обучения

Владислав Пыж

МФТИ

Цели работы

Нашей целью является прогнозирование экстремальных событий, а именно наводнений для дальнейшего использования этих данных для экономического прогнозирования

Своей задачей мы поставили прогнозирование количества осадков на ближайшие несколько дней в Калифорнии.

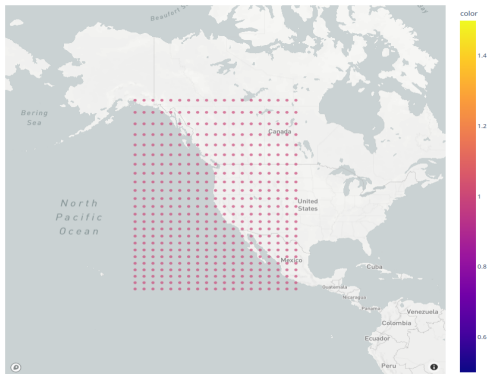
В своей работе в качестве базовой модели мы используем классические методы машинного обучения. С базовой моделью мы сравниваем результаты полученные с помощью сверточных нейронных сетей.

Литература

- ▶ Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, Xiangnan He. Modeling Extreme Events in Time Series Prediction. KDD-2019
- ▶ Roman Kail, Alexey Zaytsev, Evgeny Burnaev. Recurrent Convolutional Neural Networks help to predict location of Earthquakes.
- ▶ Nikolay Laptev, Jason Yosinski, Li Erran Li, Slawek Smyl. Time-series Extreme Event Forecasting with Neural Networks at Uber.

Описание данных

Coords CASM



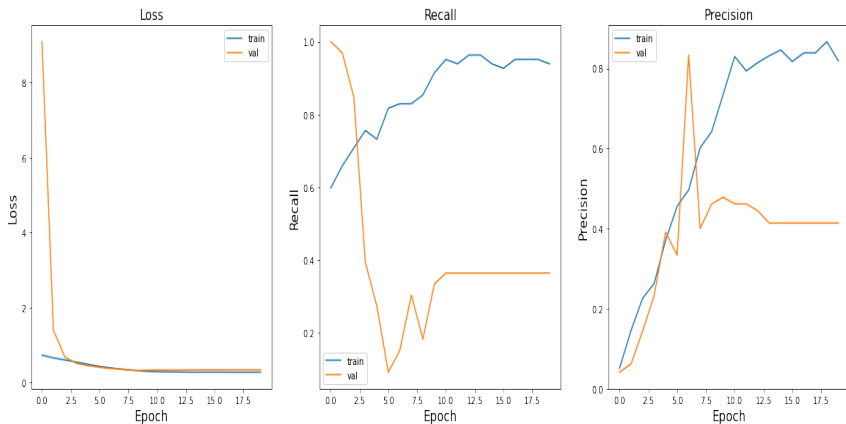
Данные с информации на территории США

Критерий модели

Мы используем в качестве критерия модели пару
рекол-пресижн

За последние 40 лет на территории Калифорнии было всего
около 120 экстремальных дней

Сходимость модели



Процесс обучения занял порядка 20 минут

Сравнение с бейзлайном

Наша модель получила результат около 0.41 пресиж и 0.36 рекол.

	accuracy	precision	recall	conf_mat	name	sampling
0	0.980044	1.000000	0.052632	[[883, 0], [18, 1]]	CatBoost	None
1	0.949002	0.170732	0.368421	[[849, 34], [12, 7]]	CatBoost	Smote
2	0.733925	0.066667	0.894737	[[645, 238], [2, 17]]	CatBoost	Under
3	0.978936	0.000000	0.000000	[[883, 0], [19, 0]]	SVC-linear	None
4	0.829268	0.090909	0.789474	[[733, 150], [4, 15]]	SVC-linear	Smote
5	0.770510	0.076577	0.894737	[[678, 205], [2, 17]]	SVC-linear	Under
6	0.978936	0.000000	0.000000	[[883, 0], [19, 0]]	SVC-rbf	None
7	0.746120	0.066116	0.842105	[[657, 226], [3, 16]]	SVC-rbf	Smote
8	0.427938	0.015595	0.421053	[[378, 505], [11, 8]]	SVC-rbf	Under
9	0.977827	0.333333	0.052632	[[881, 2], [18, 1]]	LogReg	None
10	0.833703	0.093168	0.789474	[[737, 146], [4, 15]]	LogReg	Smote
11	0.745011	0.069388	0.894737	[[655, 228], [2, 17]]	LogReg	Under
12	0.980044	1.000000	0.052632	[[883, 0], [18, 1]]	RandomForest	None
13	0.952328	0.184211	0.368421	[[852, 31], [12, 7]]	RandomForest	Smote
14	0.747228	0.069959	0.894737	[[657, 226], [2, 17]]	RandomForest	Under

Итоги

На данном этапе пока не получилось преодолеть наш бейзлайн

Готов пайплайн для изменения модели и тестирования новых гипотез