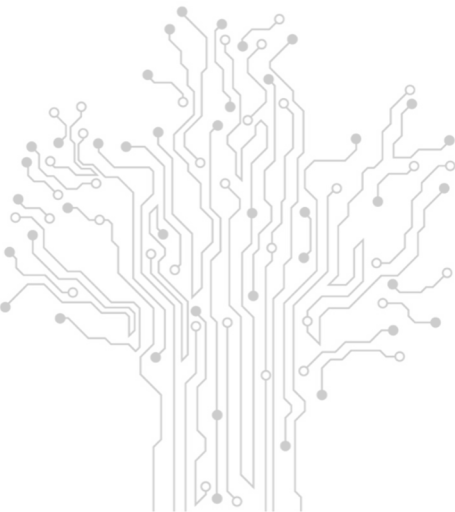


Прогнозирование показателей доходности и риска банков с помощью текущих результатов деятельности и анализа новостей

Prediction of banks' performance and risks using current business results and news sentiment analysis

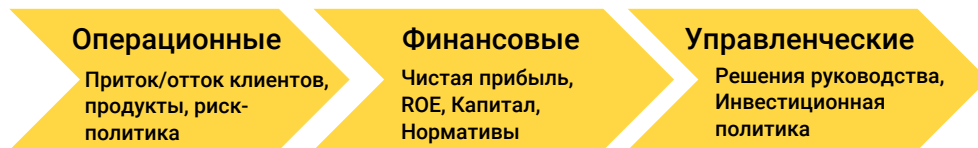
Козачкин Тимофей, МОВС 2022



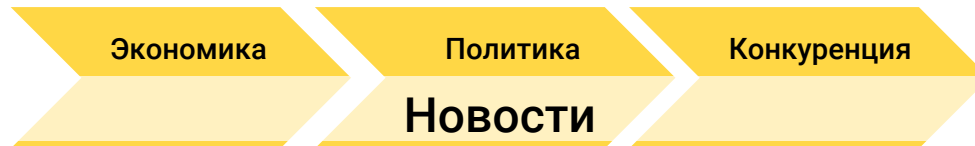
Постановка задачи

Что оказывает влияние на финансовые результаты деятельности банковской организации?

Внутренние показатели



Внешние показатели



Гипотеза:

- предсказание на основе комбинации факторов
- предсказание не только стоимости банка, а других показателей деятельности

Постановка задачи

Особенности

- Построить модель для российского рынка
- Обучить модель на русскоязычных текстах новостей
- Использовать модели и обработку текстов специфичных для русского языка



Банк России

Постановка задачи

Цель:

создать модель, комбинирующую данные о результатах деятельности банков и новостной фон вокруг них, которая будет способна предсказывать показатели доходности и риска кредитных организаций

Этапы:

1

Построить модели на основе показателей банков
(финансовая отчетность, бизнес-результаты, капитал, нормативы и пр.)
Предсказывать *вероятность дефолта и попробовать другие показатели в качестве таргета*

2

Построить модели, способные предсказывать результаты деятельности на основе текстов новостей:

- Решить задачу классификации – предсказывать направление изменений
- Решить задачу регрессии – количественно предсказывать показатели

3

Скомбинировать результаты моделей

Данные

Данные о банках

Russian banks' defaults Dataset

- >70 тыс. наблюдений
- с 2010 по 2018 гг.
- **Номер лицензии**
- **Дата**
- Сумма портфеля ценных бумаг
- Сумма инвестиций в др. организации
- Сумма кредитов
- Сумма непогашенных кредитов
- Сумма депозитов
- Изменение курса валют USDRUB
- Изменение чистых валютных активов
- Изменение финансовых обязательств банка
- Изменение стоимости привлеченного капитала
- Спред по кредитной ставке
- Размер капитала Мск/Спб
- Ставки MICEX, MIACR (МБК)
- Изменение гос. Долга
- резерв по кредитам
- Сумма залогов по кредитам
- Сумма кредитного портфеля
- Чистые активы
- Чистая прибыль
- ROA, ROE
- Нормативы N1, N2, N3
- **Событие дефолта**

Новостные данные

Фильтры для новостей:

- Период: с 2010 по 2018 гг.
- Темы: Экономика, Политика, Россия

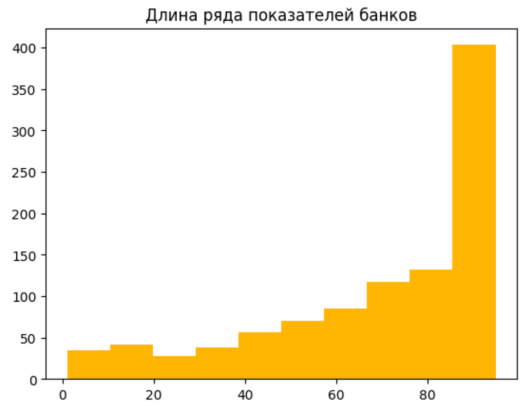
После фильтрации осталось около
400 тыс. новостей

Источники

Датасет	Количество наблюдений / Объем данных
Russian Telegram chats history	177 историй новостных чатов (3GB)
News dataset from Lenta.Ru	800k
Russian News	21k

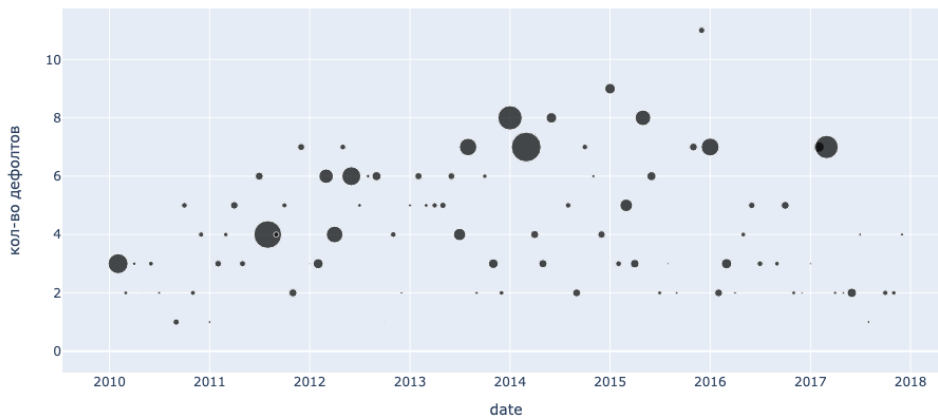
Данные (1)

- 1006 банков



- 373 событий дефолта из 70 тыс. наблюдений

Распределение дефолтов по времени



Модель прогнозирования дефолта

Catboost

Прогнозирование события дефолта для банка

METRIC	mean	pre-last value	last value
PREC_DEFAULT	1	1	1
REC_DEFAULT	0.938	0.9062	0.9687
PRE_NOT_DEFAULT	0.972	0.9583	0.9857
REC_NOT_DEFAULT	1	1	1
F1_DEFAULT	0.968	0.9508	0.9841
F1_NOT_DEFAULT	0.986	0.9787	0.9928
F1_ALL	0.977	0.9645	0.9884
NOT_DEFAULT	69	69	69
DEFAULT	32	32	32
NOT_as_NOT	69	69	69
NOT_as_DEF	0	0	0
DEF_as_NOT	2	3	1
DEF_as_DEF	30	29	31
ACCURACY	0.98	0.9702	0.99

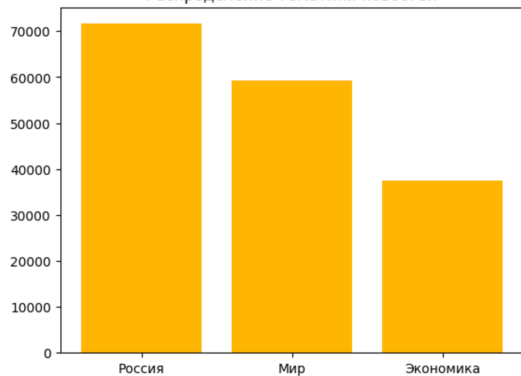
Catboost

Прогнозирование периода дефолта

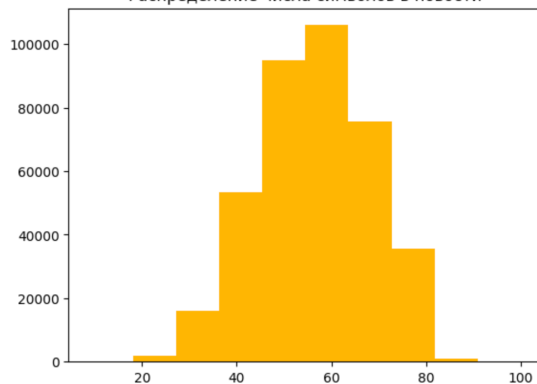
METRIC	last value
PREC_DEFAULT	1
REC_DEFAULT	0.5675
PRE_NOT_DEFAULT	0.9977
REC_NOT_DEFAULT	1
F1_DEFAULT	0.7241
F1_NOT_DEFAULT	0.9988
F1_ALL	0.8395
NOT_DEFAULT	7012
DEFAULT	37
NOT_as_NOT	7012
NOT_as_DEF	0
DEF_as_NOT	16
DEF_as_DEF	21
ACCURACY	0.9977

Данные (2)

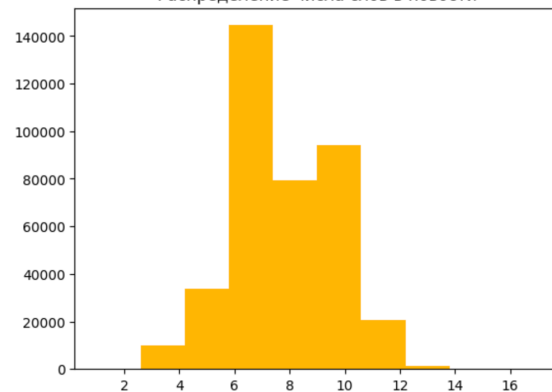
Распределение тематики новостей



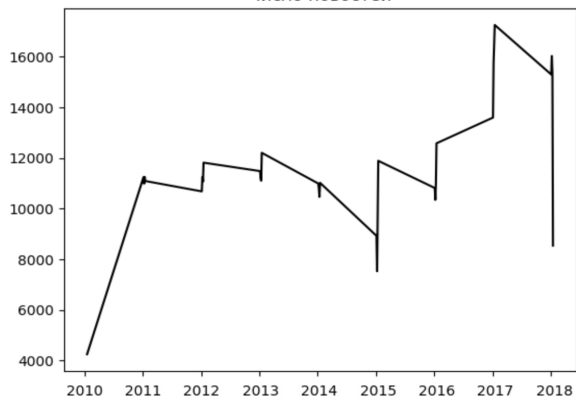
Распределение числа символов в новости



Распределение числа слов в новости



Число новостей

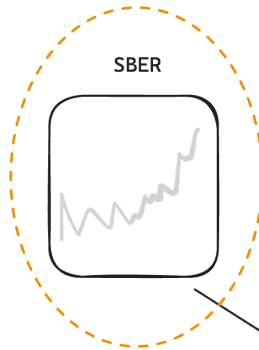


Задача прогнозирования по новостям



Датасет для обучения

1. for one stock



News

date	news

2. for many indicators

bank1	bank2	bank3
ROE	ROE	ROE
NI	NI	NI
N1,2,3	N1,2,3	N1,2,3
Capital	Capital	Capital

date	price/indicator	news

sentiment features and/or embeddings

Обработка текста

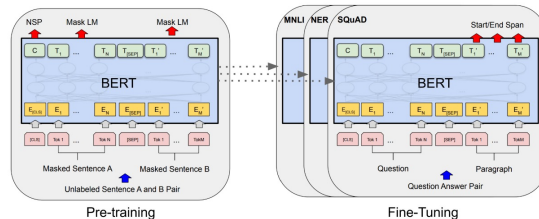
Tokenizer

- Lemmatizing text



- TF-IDF
- Word2Vec

Sentiment Analysis



- FinBERT
- sbert-ru-sentiment-rusentiment

*Fine-Tuning на основе новостей.
Предсказываем score по классам:
0 – negative
1 – neutral
2 – positive*

LSTM сеть из 5 слоев:

- 1 input, 3 hidden, 1 output.
- shape of the input (10, 1)
- "tanh" activation function.
- 1 neuron in the output layer
- 3 dropout layers (with dropout rate 0.1, 0.05, 0.01 respectively)
- MSE loss function
- Adam optimizer having 0.02 learning rate.

- 10-day rolling window
- (11, 1) if we add sentiment score

Модель прогнозирования дефолта

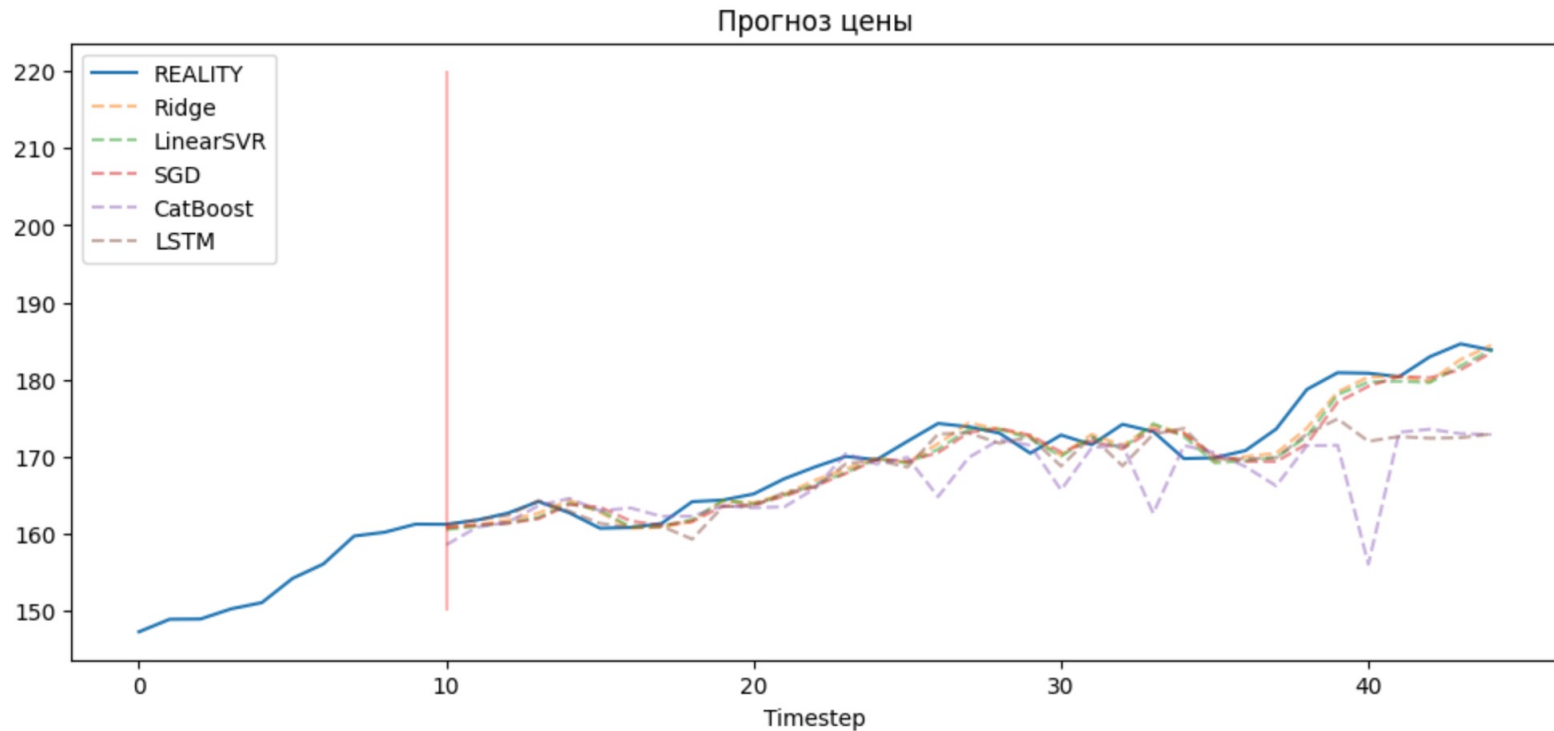
Классификация

Accuracy	without sentiment			with sentiment		
	without tokenizer	TF-IDF	Word2Vec	without tokenizer	TF-IDF	Word2Vec
LogisticRegression	0.5544	0.5777	0.5751	0.5783	0.5892	0.5901
SVC	0.5647	0.5855	0.5855	0.5702	0.5923	0.5891
SGD	0.5259	0.5906	0.5932	0.5301	0.6193	0.6284
Catboost	0.52	0.5207	0.5311	0.5456	0.5568	0.551
LGBM	0.5829	0.528	0.5207	0.5875	0.5567	0.5647

Регрессия

RMSE	without sentiment			with sentiment		
	without tokenizer	TF-IDF	Word2Vec	without tokenizer	TF-IDF	Word2Vec
Ridge	2.49	2.39	2.46	2.44	2.38	2.354
LinearSVR	2.61	2.37	2.84	2.53	2.234	2.932
SGD	2.59	32.8	3.08	2.65	2.958	2.85
Catboost	2.48	2.66	2.88	2.38	2.54	2.45
LGBM	3.05	2.87	2.61	2.93	2.75	2.745
LSTM Architecture	2.04			1.92		

Модель прогнозирования дефолта



Сервис



18:20 Локо-Банк повысил ставки по кредитам

👁 65

17:57 В Кабмине назвали продукт, который станет альтернативой ипотеке

👁 212

17:24 В ЦБ раскрыли, на сколько еще собирались поднять ключевую ставку

👁 1158

16:41 Все средства хороши: объяснено, зачем россияне массово кинулись брать кредиты

👁 919

16:31 ЦБ установил официальные курсы валют на 14 мая

👁 387

15:58 Когда следует купить мини-каско: советуют эксперты

👁 300

15:55 Аналитики дали совет по вкладам и накопительным счетам

👁 4153

News text to insert

bank name

Показатель	Прогноз
stocks	— ...
ROE	↗ ...
ROA	↖ ...
Net Income	↖ ...
N1	↗ ...
N2	↗ ...

- PostgreSQL
(для хранения информации о банках)
- Flask, Gunicorn
(для сервера)

Заключение

1

Добавить данных:

- новостей, содержащих имена банков, а не только на общие темы

2

Попробовать другие эмбединги

(*BertTokenizer - rubert-base-cased-sentiment, sbert-ru-sentiment-rusentiment*)

3

Поэкспериментировать с таргетами, попробовать предсказать не только стоимость акции

4

Скомбинировать результаты 2х моделей для прогноза дефолта/показателей

Спасибо за внимание!