



# Билибин Сергей Константинович

Мужчина, 27 лет, родился 13 июля 1995

+7 (911) 9400158 — предпочитаемый способ связи

s.bilibin2017@yandex.ru

Другой сайт: <https://t.me/bilibinsk>

Проживает: Санкт-Петербург

Гражданство: Россия, есть разрешение на работу: Россия

Готов к переезду: Москва, не готов к командировкам

## Желаемая должность и зарплата

**data scientist/аналитик**

**50 000**  
руб.

Специализации:

— Аналитик

Занятость: полная занятость

График работы: полный день

Желательное время в пути до работы: не имеет значения

## Опыт работы — 2 года 5 месяцев

Июнь 2021 —  
Июль 2022  
1 год 2 месяца

### Webmasters

#### data scientist

Компания занимается анализом и прогнозированием результатов киберспортивных событий (dota2, league of legends, csgo).

Моя задача – создать инфраструктуру для анализа и прогнозирования турнирных игр по cs:go.

Задача включала в себя следующие подзадачи:

1. Создать вспомогательную систему для аналитиков, позволяющую проводить анализ событий
2. Создать модель машинного обучения, прогнозирующую исходы событий

МЛ пайплайн состоял из следующих этапов:

1. Парсинг источника данных: API. Парсеры написаны на python, работали по cron: регулярно заполняли MySQL БД.
2. БД состояла из 11 таблиц: лига, серия, турнир, матч, игра, статистика, раунды, карта, игрок, команда, игроки для команд. Данная БД использовалась для создания признаков и для создания дашборда о предстоящей игре (streamlit+fast\_api)
3. Очистка данных(ошибки, аномалии, пропуски)
4. Создание признаков:
  - 4.1. профайлинг статистики истории игр игроков и команд
  - 4.2. обучение эмбедингов каждого игрока и команды
  - 4.3. паспорт игрока и команды: город, страна, национальность, возраст
  - 4.4. карта, на которой проводится игра, сторона начала для команд (террористы/контр-террористы), турнирный этап(группа, плей-офф, финал и др), тир игры (регионы имеют разный уровень игры-всего 5 уровней)
  - 4.5. Собирались данные об активности игроков в социальных сетях
  - 4.6. Собирались результаты игр игроков в рамках матчмейкинга (обычные ранговые игры игроков вне турниров).
5. Было сгенерировано около 100 000 признаков. Отбор осуществлялся по следующей схеме:

- 5.1. Признаки бились на группы по 1000 шт.
- 5.2. В каждой группе с помощью ранней остановки обучалась модель. С помощью permutation importance(изменения метрики при перемешивании) отбирались важные признаки
- 6.Отобранные и выброшенные признаки формировали 2 датасета(выброшенные признаки ранжировались и последовательно добавлялись в отобранный датасет)
7. Оптимизировались гиперпараметры модели (lightgbm, catboost). Повторно производился отбор признаков
8. Производилась аугментация признаков (менялись местами колонки с характеристиками игроков одной команды команды)
9. Производился стекинг
10. Деплой проводился с помощью FastAPI+Streamlit+Docker

#### РЕЗУЛЬТАТ:

С помощью приложения аналитик мог провести статистический анализ предстоящей игры,оценить шансы различных исходов, сделать прогноз, анализировать результаты прогнозов.

Июль 2019 —  
Сентябрь 2020  
1 год 3 месяца

## БК Балтбет

### Аналитик

- 1) написание регулярных отчетов об игроках, видах спорта и тд., мониторинг событий (ежедневный отчет об убытках:причина, сумма; выявление закономерностей)
- 2) выполнение проектов
  - 2.1.) реализовал ранжирование пользователей
  - 2.2.) реализовал сегментирование относительно характера поведения
  - 2.3.) реализовал модель оттока пользователей

## Образование

### Бакалавр

2020

#### Санкт-Петербургский государственный университет экономики и финансов, Санкт-Петербург

прикладная математика, Математика. Прикладная математика

2018

#### Санкт-Петербургский государственный университет экономики и финансов, Санкт-Петербург

математические методы в экономике, математические методы в экономике

## Ключевые навыки

Знание языков

Русский — Родной  
Английский — C1 — Продвинутый

Навыки

Обучаемость

## Дополнительная информация

Обо мне

Закончил физ-мат лицей, поступил на «мат. методы в экономике» в СПбГЭУ. В бакалавриате занимался оптимизацией, проходил практику в «Невафильм» - писал диплом о размещении

сервисных центров для кинотеатров. Продолжил обучение в магистратуре на направлении «Прикладная математика и информатика». Увлёкся машинным обучением и писал диплом на тему «Анализ рынка углеводородов»: применял различные методы машинного обучения для прогнозирования временных рядов. Прошел курс ODS(Open Data Science), принимал участие в соревнованиях на kaggle. После обучения работал в БК БАЛТБЕТ(букмекерская контора): провел сегментирование пользователей и создал модель оттока(отточным пользователям предлагались бонусы в соответствии с их предпочтениями) . Отработал 1 год и взял перерыв на прохождение Яндекс.Практикума «Data science»: все темы для меня уже были известны, но главный итог – это прогресс в постановке задач, совмещении бизнес целей и мат. моделей, оформлении и презентации решений – каждая работа проверялась и возвращалась на доработку с комментариями. После окончания курсов (июнь 2021) устроился в IT стартап (Webmasters), занимающийся разработкой сервисов для киберспортивных событий, на позицию data scientist. Разработкой пока что занимаюсь только я, поэтому участвую на всех этапах разработки: взаимодействие с бизнесом(понимание задачи) -> постановка мл задачи -> сбор данных -> первичный анализ -> очистка данных -> выдвижение гипотез -> генерирование признаков ->валидация моделей -> финальная модель -> создание пользовательского приложения для использования модели

Стек инструментов:

1. Python
2. Sql(my/postgre/clickhouse)
3. Apache Spark, Hadoop, Bigquery
4. PowerBI
5. Git
6. Docker
7. Flask
8. FastAPI
9. Streamlit
10. Dash
11. Pandas
12. Numpy
13. Scipy
14. Statsmodels
15. Sklearn
16. XGBoost
17. LightGBM
18. Keras
19. PyTorch
20. NLTK
21. Работа в командной строке Linux OS/Windows OS
22. Составление расписания скриптов с помощью Cron
23. Есть опыт постановки мл задачи на основе потребностей бизнеса, предоставления результатов исследования с целями, задачами, средствами, результатами исследований; есть опыт совместной разработки с командой из Сколково
24. резидент Сколково
25. провожу собеседования при подборе дата саентистов
26. работаю в YouTracker/Jira. Ставится задача, обсуждается с командой, делится на спринты, выполняется задача, делается код-ревью, операция повторяется.
27. есть опыт работы на сервере
28. есть опыт деплоя
29. есть опыт многопоточной и асинхронной разработки (multiprocessing, threading, asyncio)

Диплом о профессиональной переподготовке Яндекс.Практикум: специализация "Data science"

[https://github.com/sergbill/yandex\\_praktikum\\_ds\\_projects/blob/main/Сергей%20Билибин\\_20212DS00434.pdf](https://github.com/sergbill/yandex_praktikum_ds_projects/blob/main/Сергей%20Билибин_20212DS00434.pdf)

Архив с портфолио, включающий:

- 1) диплом yandex практикум "data science"
- 2) проекты yandex практикум "data science"
- 3) соревнования kaggle
- 4) парсер hh.ru (подбор вакансий)
- 5) парсер youtube-канала (статистика канала)

ССЫЛКА(кнопка "download" в правом верхнем углу):

[https://github.com/sergbill/yandex\\_praktikum\\_ds\\_projects/blob/main/portfolio.rar](https://github.com/sergbill/yandex_praktikum_ds_projects/blob/main/portfolio.rar)