

АНАЛИТИКА МОБИЛЬНОГО МАРКЕТИНГА В ПРИЛОЖЕНИЯХ: КАК ПРИ СУЩЕСТВУЮЩЕМ НАБОРЕ ИНСТРУМЕНТОВ ПРОВОДИТЬ СБОР ДАННЫХ

Владислав Денисов



SPORTS.RU

Спортс – digital-издательство для неравнодушных к спорту

MAU

30 млн

уникальных пользователей

DAU

1,5 млн

уникальных пользователей

SOCIAL MEDIA

10 млн

уникальных пользователей

У нас огромное число UGC-контента, мы снимаем фильмы, создаем игры и много чего еще



здесь подробнее про нас

КОМАНДА

- 4 «data fullstack» человека
- Нет разделения по ролям (веб/мобайл, продукт/маркетинг)
- Есть условное разделение по зонам ответственности (описание, инжиниринг, nlp, исследования, спорт)
- Нанимали людей даже без знания SQL, главное – математика
- Аналитик – не калькулятор и не мастер по презентациям: все отчеты автоматизированы

ДААННЫЕ В SPORTS.RU

Ключевые для бизнес-логики

Спортивная статистика
(~ 3к событий на матч)

Новости и материалы (~ 3к в месяц)

Активность пользователей:
комментарии (до 80к в сутки),
рейтингование (до 250к в сутки)
и подписки

Картинки
и другие медиа-файлы

ДААННЫЕ В SPORTS.RU

Аналитические

Кликстрим: ежедневно +50 ГБ
сжатых данных о взаимодействии
пользователей с сайтами,
приложениями и другими сервисами

Экспорты из разных платформ
(реклама, CRM, технический
мониторинг и т.д.) для обогащения
отчетов

ДААННЫЕ В SPORTS.RU

Приложения

Старый формат: все события отправляются в Appmetrica, дальше выгружаются через Logs API к нам

Новый формат: данные отправляются напрямую к нам, но сохраняем связь с идентификаторами Appmetrica

АНАЛИТИКА

Инструменты



AppMetrica

- До внедрения платных сервисов использовали как платформу для сбора ивентов и дальнейшей выгрузки.
- Сейчас используем для отправки push-кампаний по сегментам аудитории и трекинга источников



Firebase

- Remote Config – инструмент для управления настройками приложения с сервера
- A/B-эксперименты – только для создания групп, результаты считали сами

АНАЛИТИКА

Свой счётчик. Зачем?

- Возможность построения кастомных метрик и отчетов
- Мощная детализация (фильтры по конкретному пользователю или последовательности событий)
- Единая структура для веба и приложений
- Соединение с метаданными для красивых дашбордов
- Открыт в open-source, готовы консультировать по внедрению / использованию

АНАЛИТИКА

Базы данных



АНАЛИТИКА

ClickHouse

- Колоночная СУБД, изначально разработанная Яндексом (сейчас отдельная компания), хранилище для нашего кликстрима и других сервисов, где важны не единичные строки, а агрегаты.
- Используем с 2016-го года, тогда – удачная находка (скорость работы позволяла закрыть глаза на некоторые баги и непривычный синтаксис и бросать все силы на переезд с Redshift), сейчас – один из лидеров в области.
- Благодаря шардированию (параллельное вычисление и хранение данных) и репликации (дублирование данных между серверами для отказоустойчивости) можно не терять в скорости с увеличением объемов, данных, в отличие от классических версий MySQL / PostgreSQL.

АНАЛИТИКА

Инструменты



kubernetes

FastAPI




Apache
Airflow



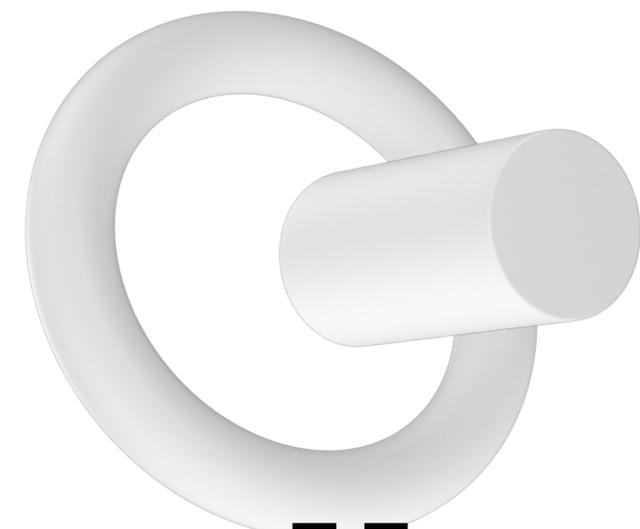
re dash



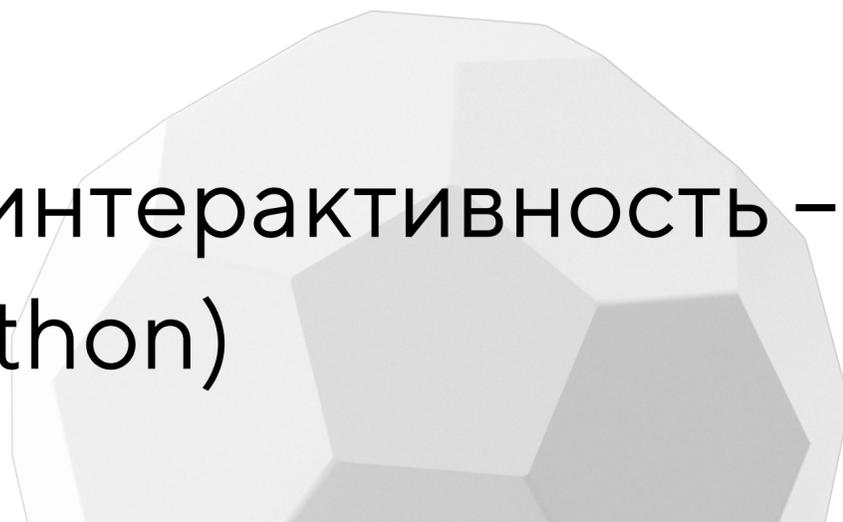
Streamlit

АНАЛИТИКА

Визуализация



- Минусы своего счетчика – на выходе просто таблица в БД, все представления данных нужно собирать с нуля
- BI – Redash (даже контрибьютим в развитие) – огромный набор визуализаций и возможность собирать дашборды на Python
- Если нужна интерактивность – пишем приложение на Streamlit (Python)



Date

13/11/22 → 16/11/22 ⚡

Tournament

world-cup [Dropdown]

Source_type

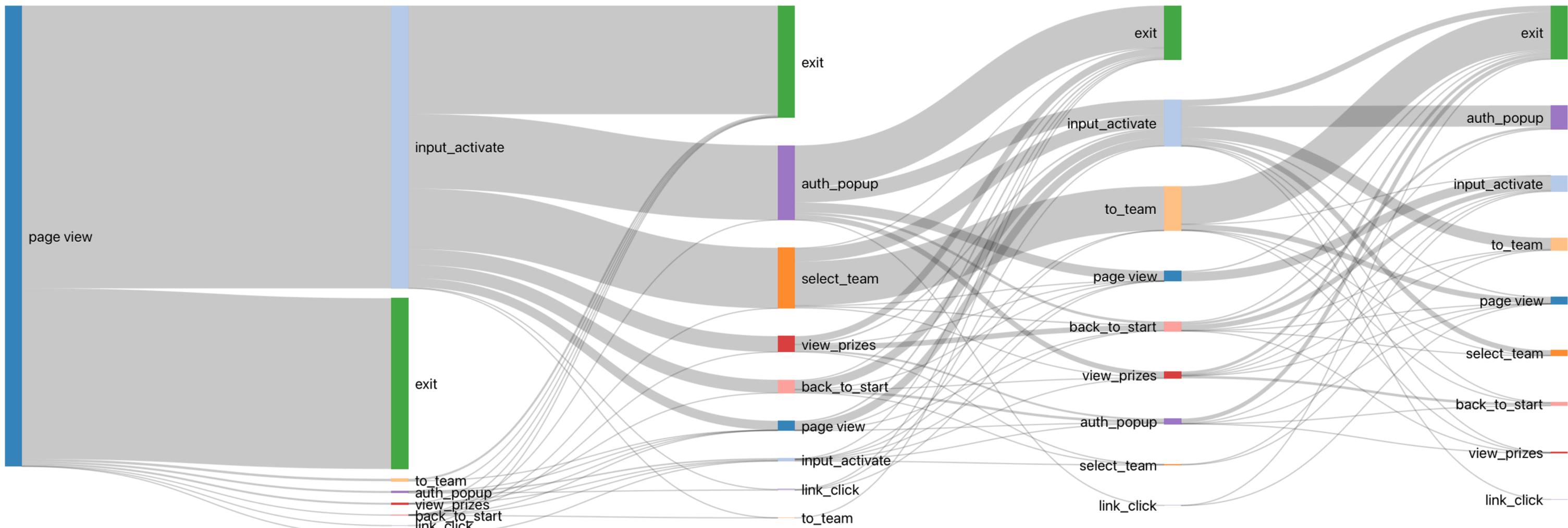
unknown [Dropdown]

Fantasy | Onboarding | Sankey

Описание: <https://docs.google.com/spreadsheets/d/1ZwiQReLj00TXefNgPt3YMVYXmMoHWGM3IZw1HUa-Tcc/edit#gid=2131333336&range=D141:D148>

Tournament

world-cup [Dropdown]



АНАЛИТИКА

Конкретнее про бигдату и детализацию

- Суммарно около 1000 строк в таблице с описанием событий
- Отслеживаем просмотр страницы и явные и неявные действия пользователей на них, каждое действие – 1 строка в базе, а 1 просмотр страницы генерирует около 100 событий
- Примеры вопросов, на которые умеем отвечать: какая дочитываемость у контента? а только у тех, кто заходит 5 дней подряд и чаще всего читает про Спартак? сколько денег нам приносит пользователь из ВК? какое пересечение у пользователей между вебom и приложениями? что мешает пользователям в процессе регистрации?



АНАЛИТИКА

Data Discovery

- Знания по структуре БД и событий не должны лежать в людях – выносим в общедоступные ресурсы и стараемся поддерживать актуальность
- Документация ивентов – в Google Sheets
- Документация БД – в DataHub



Google
Sheets



DataHub

СЕРВИСЫ НА ОСНОВЕ ДАННЫХ

- Аналитики в Sports.ru почти не занимаются ресерчем в ML: часто используем лучшие практики в уже готовых библиотеках
- Команда аналитики сама пишет backend для таких сервисов
- Нейросети – не панацея, если классические алгоритмы справляются не хуже, чаще «классика» работает значительно быстрее, что важно для продакшна
- Есть задачи, которые приходят от бизнеса, но значимая доля сервисов рождается из желания попробовать новый инструмент

СЕРВИСЫ НА ОСНОВЕ ДАННЫХ

Сегментация пользователей

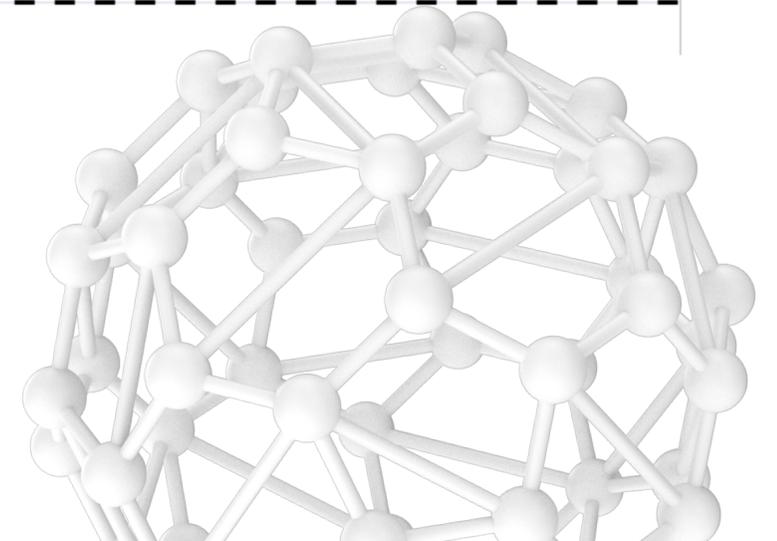
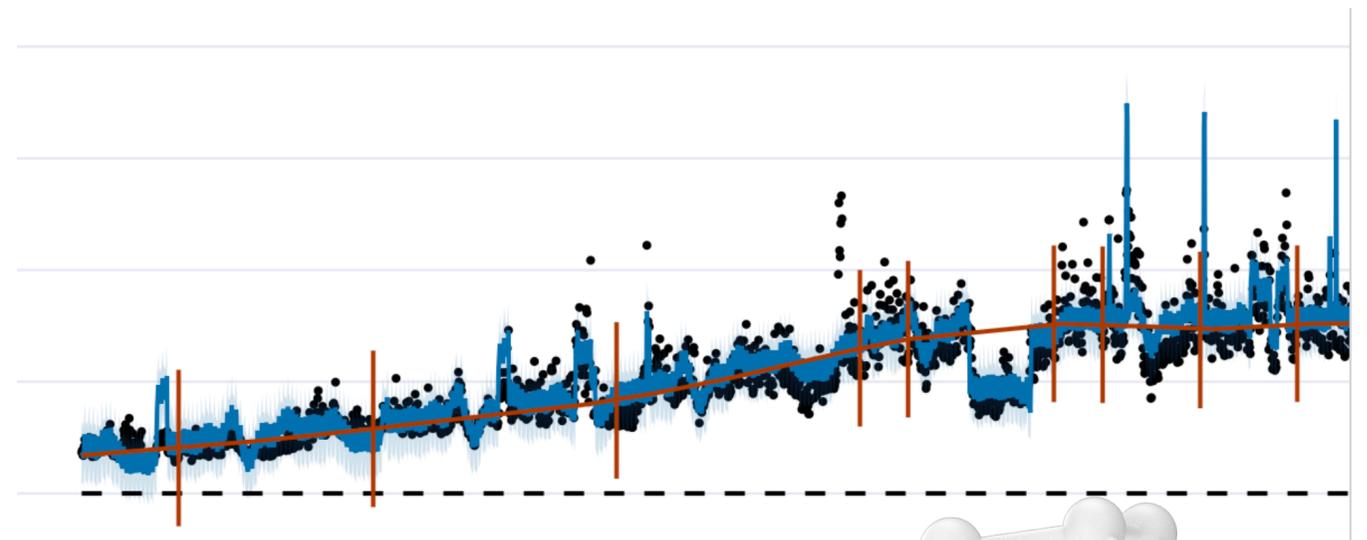
- Разделение пользователей на сегменты: по видам спорта, характеру потребления и взаимодействия с сервисами
- Определяем на своей стороне, дальше отправляется событие о принадлежности к группам в любую систему аналитики, маркетинга или рекламы
- Повышаем качество маркетинга, пытаюсь найти похожих пользователей



СЕРВИСЫ НА ОСНОВЕ ДАННЫХ

Прогноз трафика

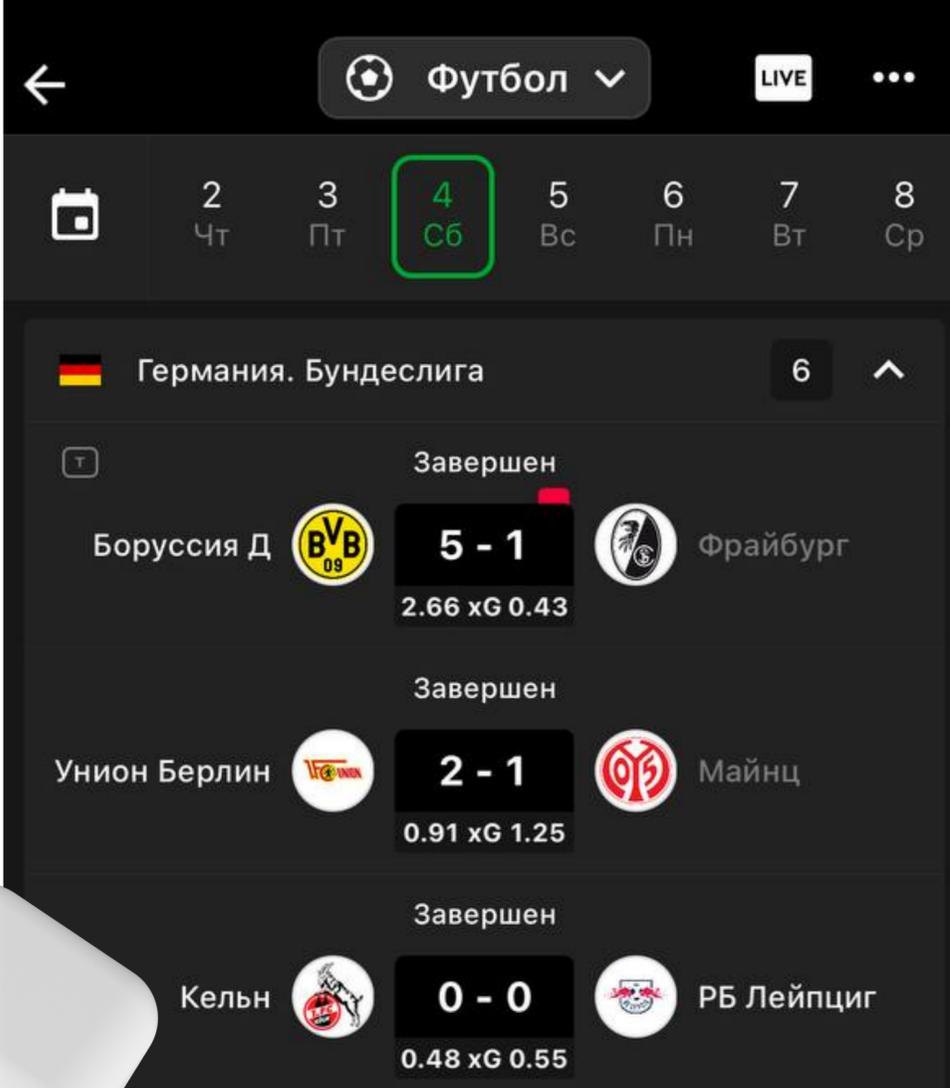
- У нас рекламная модель бизнеса – выручка коррелирует с просмотрами страниц
- Делаем краткосрочный (месяц) и долгосрочный (год) прогнозы
- Сервис прогноза – инструмент для отдела продаж – делаем так, чтобы рекламы было ровно столько, сколько инвентаря доступно



СЕРВИСЫ НА ОСНОВЕ ДАННЫХ

Expected goals в футболе

- xG – показатель расширенной футбольной статистики, который показывает остроту моментов
- Обучили модель классификации на основе бустинга, которая предсказывает вероятность гола по 100+ параметрам
- Топ-5 лиги + РПЛ имеют данный показатель, который обновляется прямо во время матча



The screenshot shows a mobile application interface for football. At the top, there is a navigation bar with a back arrow, a 'Футбол' (Football) dropdown menu, a 'LIVE' indicator, and a menu icon. Below this is a calendar view with days 2 (Чт), 3 (Пт), 4 (Сб), 5 (Вс), 6 (Пн), 7 (Вт), and 8 (Ср). The date '4 Сб' is highlighted with a green box. The main content area displays the 'Германия. Бундеслига' (Germany. Bundesliga) league. Three matches are listed, each with the status 'Завершен' (Completed), the team names, the score, and the xG values for both teams.

Match	Score	Home xG	Away xG
Боруссия Д vs Фрайбург	5 - 1	2.66	0.43
Унион Берлин vs Майнц	2 - 1	0.91	1.25
Кельн vs РБ Лейпциг	0 - 0	0.48	0.55

СЕРВИСЫ НА ОСНОВЕ ДАННЫХ

Кросспроектные рекомендации

- Делаем свою систему рекомендаций с 2017-го года, сейчас подключаем сторонние ресурсы (пока бесплатный SaaS)
- Работаем качественнее «обменок» (CTR + глубина), но без кликбейта
- Множество алгоритмов: многорукие бандиты, «классика», item2item, next-item – подстраиваемся под конкретную задачу

Рекомендуем

Энцо в «Челси» за 121 млн, Забитцер и Кейлор теперь в АПЛ. Европа...

+109 70

«Челси» утащил Энцо Фернандеса! Главное о невероятной зимней закупке

+67 131

«Лю, не т... бы»: пиц...

ВСЕ!

Остались вопросы? Хотите потестировать
рекомендалку или счетчик аналитики? Пишите!

TELEGRAM: @VLADENISOV

MAIL: DENISOV@SPORTS.RU

