



## Использование Data Science для создания сервисов в Cropwise

Павел Сазонов, Data Science Lead, Syngenta



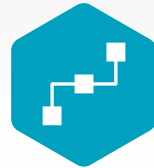
# ВВЕДЕНИЕ В DATA SCIENCE

**DATA SCIENCE** — Это область, которая использует статистические методы, программирование, анализ данных и машинное обучение для извлечения знаний и паттернов из структурированных и неструктурированных данных



Внутренние системы

*(Битрикс24, 1С, базы данных, цифровые агро-платформы)*



Данные Телематики



Спутниковые снимки



Агроскаутинг



Данные о погоде



Данные по севообороту  
и урожайности



Отчеты агрономов



И т.д.

# ИСПОЛЬЗОВАНИЕ DATA SCIENCE

---

## С помощью Data Science можно:

- Прогнозировать сезонную и предсезонную урожайность
- Определять оптимальные даты сева, опрыскивания и сбора урожая
- Рассчитывать количество и типы удобрений для максимизации урожая
- Выявлять заболевания, сорняки, вредителей, грибковые заболевания задолго до их появления
- Определять площадь, культуру и урожайность по всей России на основании спутниковых снимков
- И многое другое...

Согласно прогнозу крупнейшей консалтинговой компании «McKinsey&Company»– потенциал от внедрения и использования технологий больших данных в сельском хозяйстве **оценивается в 486 миллиардов долларов** дополнительной прибыли по всему миру\*

\* <https://www.mckinsey.com/featured-insights/artificial-intelligence/visualizing-the-uses-and-potential-impact-of-ai-and-other-analytics>

# ОБОГАЩЕНИЕ ДАННЫХ

---

**Обогащение данных** — это насыщение информацией. То есть, процесс дополнения сырых данных той информацией, которая в исходном виде в них отсутствует, но необходима для принятия оптимальных решений.

Сингента располагает данными **66000** локальных метеорологических станций, **3** спутниковых провайдера, **17** радиолокационных станций, а также большого количества глобальных источников данных

**Суммарно – 142 уникальных показателя за последние 20 лет по всей России.**



# ДОПОЛНЕНИЕ ЛОКАЛЬНЫХ МЕТЕОСТАНЦИЙ ИСПОЛЬЗОВАНИЕМ УДАЛЕННЫХ ДАННЫХ И МОДЕЛЕЙ

Использование данных Сингенты позволяет покрыть потребность в данных с локальных метеостанций:

- Данные погоды
- Здоровье растений
- Характеристики почвы
- Землепользование
- NDVI
- Топография
- Солнечная радиация
- Относительная влажность и осадки

Перед использованием инструмента  
все данные проверяются на основании  
данных локальных метеостанций  
в любой точке России.

# МОДЕЛИ ПРОГНОЗИРОВАНИЯ ОПТИМАЛЬНОЙ ДАТЫ СЕВА

Модель базируется на почасовой проверке следующих индикаторов:



Температура



Температура почвы



Осадки за последние  
48 часов



Вероятность  
выпадения осадков



Влажность почвы

Рекомендательная система для выбора оптимальной даты сева.

Доступно для: кукурузы, яровой пшеницы, ячменя и рапса, озимой пшеницы, подсолнечника, сои.

# МОДЕЛИ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ

## ПРОГНОЗ УРОЖАЙНОСТИ



Модель прогнозирует величину урожайности, которую можно получить при условиях близких к идеальным. Прогноз используется как оценка потенциала поля и инструмент диверсификации полей

## ПРОГНОЗ СТАДИЙ РОСТА

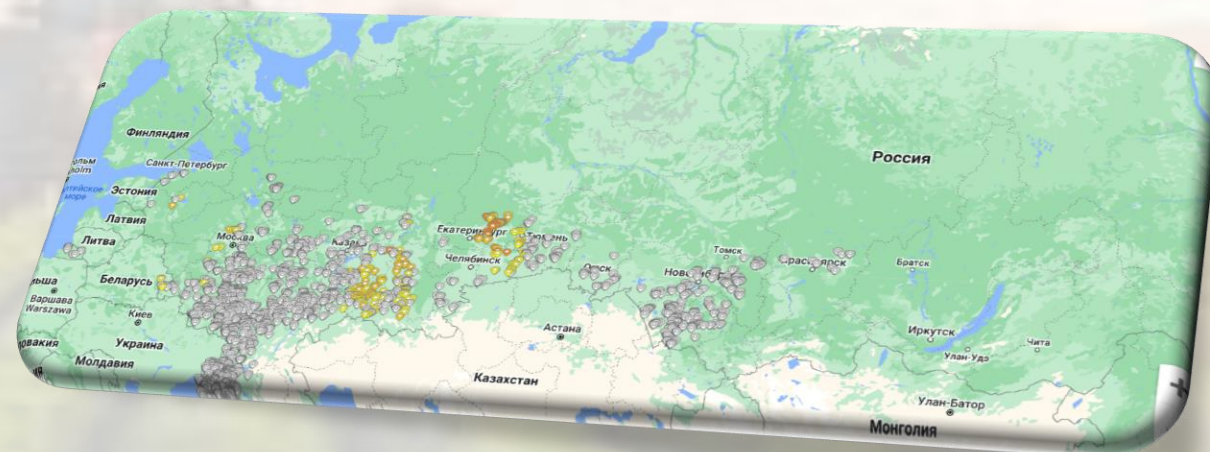







Модель прогнозирует дату наступления вегетативных фаз роста до конца сезона. Модель отличается высокой точностью и требует минимальный сет входной информации от клиента

# ИНТЕРАКТИВНАЯ КАРТА ПРОГНОЗИРОВАНИЯ ЗАБОЛЕВАНИЙ ПО ВСЕЙ РОССИИ

**Интерактивная карта прогнозирования заболеваний** позволяет выявить потенциальные очаги распространения болезней растений и превентивно использовать средства защиты растений для повышения урожайности.

**Риск рассчитывается на основании:** температуры, относительной влажности, количества осадков и влажности листьев. Все пороговые значения берут из академической литературы и при желании могут быть дополнены на основании экспертизы самого холдинга.



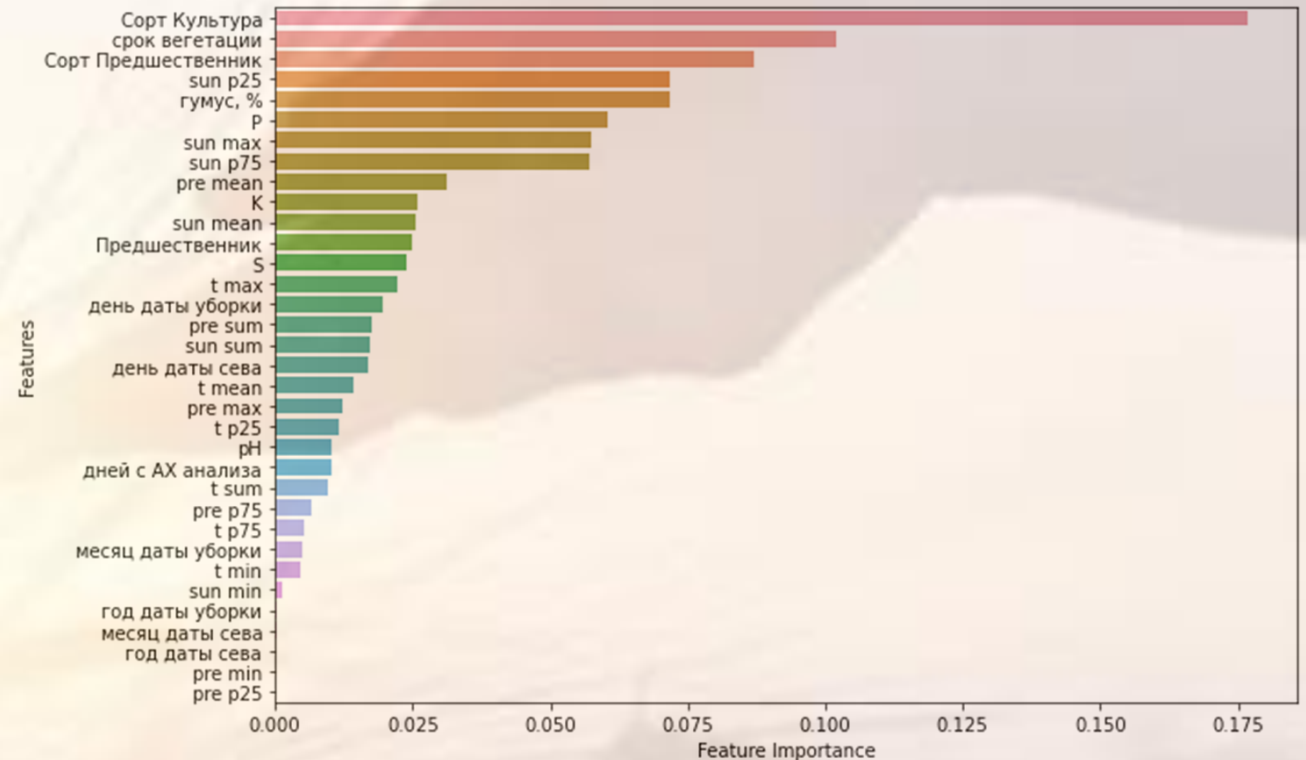
-  Отсутствует/низкий риск
-  Некоторый риск
-  Средний риск
-  Высокий риск
-  Высокий риск



# AGRIMATRIX

Цель проекта заключается в разработке модели машинного обучения, которая способна декомпозировать фактические значения урожайности на влияющие факторы и определить их веса, что позволит объяснить полученные результаты

Планируется использовать все доступные данные холдингов, выходные данные других моделей, виртуальные данные, прогнозы стадий роста, болезней, вредителей и др.



**СПАСИБО ЗА ВАШЕ ВНИМАНИЕ!**

---

Thank  
YOU