

# ПРОГНОЗИРОВАНИЕ КАЧЕСТВА ТАРНОГО КАРТОНА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Д.Н. Жирнов<sup>1</sup>, Е.В. Дернова<sup>2</sup>, Д.А. Дулькин<sup>2</sup>

*ООО «Сухонский КБК»*

*ООО «УК «ОБФ»*



**СКБК**

Сухонский  
картонно-бумажный  
комбинат

# Что же такое Big Data?

**BIG DATA** - обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами.

Характеризуются:

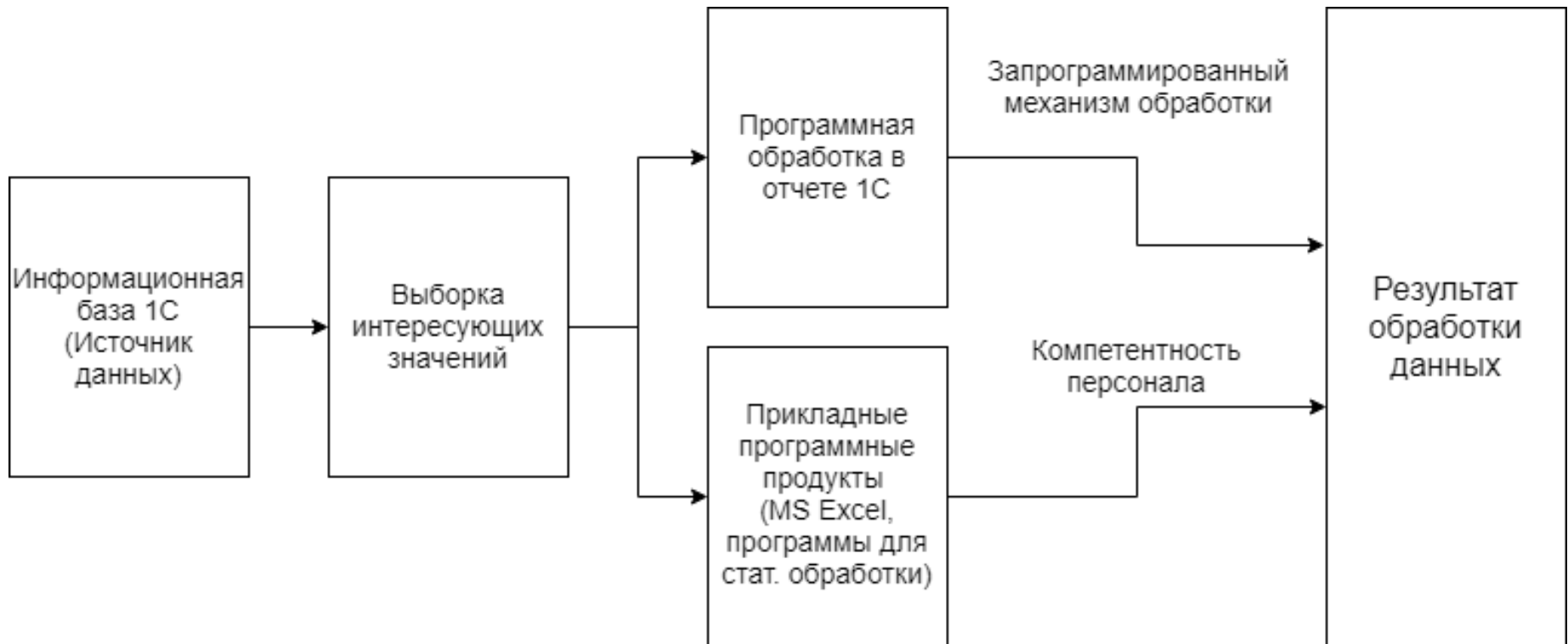
- Объемом (на физическом носителе)
- Скоростью (увеличение скорости прироста объемов информации)
- Многообразием - одновременная обработка различных типов данных.

# Анализ применимости термина «BIG DATA» к базам данных «УК «ОБФ»

В качестве объекта для изучения предлагается рассмотреть информационную базу данных 1С предприятия:

- Объем – более 500 Гб;
- Скорость изменения данных – каждую единицу времени в существующую информационную базу вносятся новые элементы или происходит изменение существующих элементов.
- Многообразие типов значений – в базе данных хранятся как значения показателей качества готовой продукции по каждому рулону, так и техническая информация об оборудовании, их обслуживанию и др.

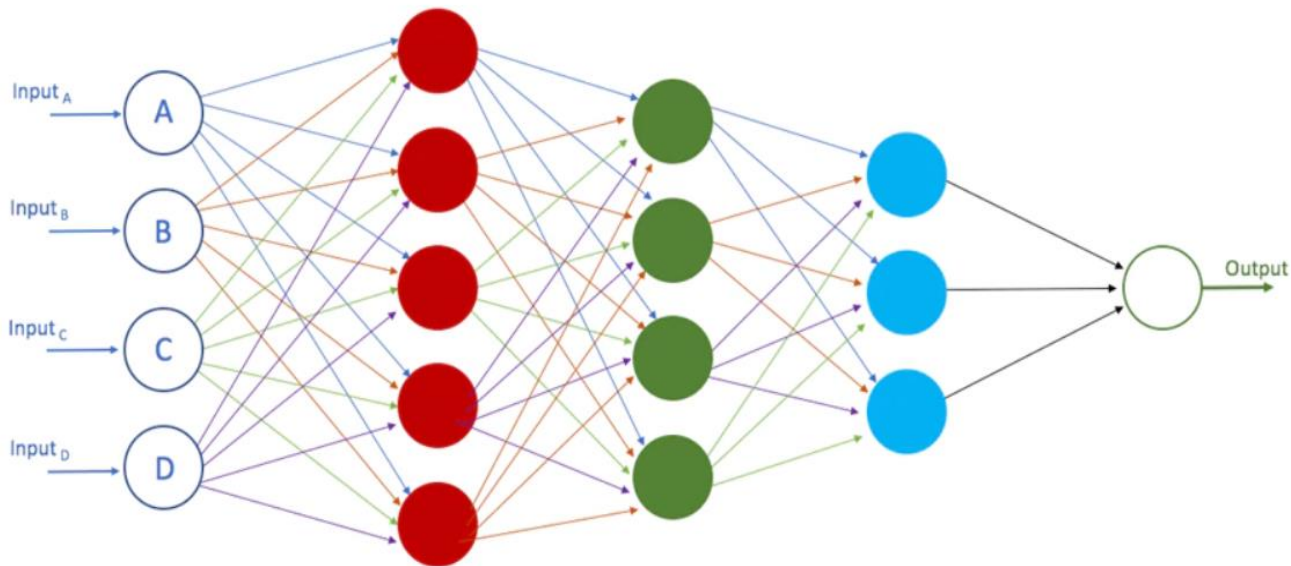
# Существующие методы обработки данных



# Использование методов обработки BIG DATA

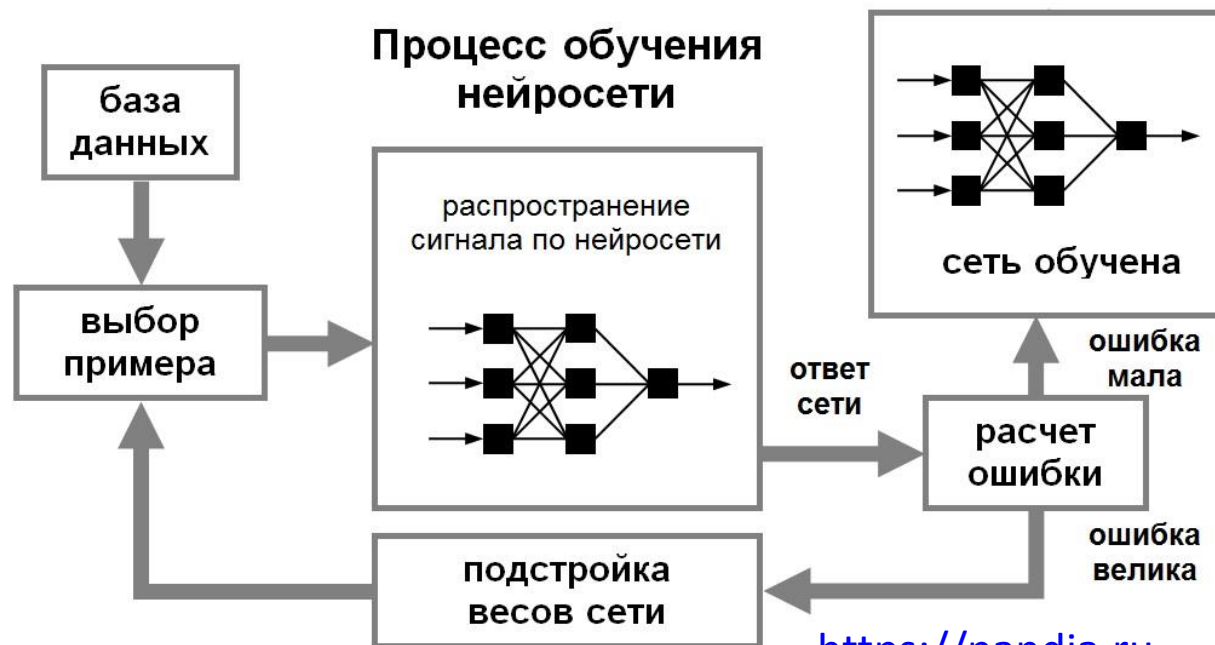
Среди многообразия инструментов обработки, с технологической точки зрения представляет особый интерес получение комплексных прогнозов.

Данные системы в основаны на методах машинного обучения - использование моделей, построенных на базе статистического анализа или машинного обучения для получения комплексных прогнозов на основе базовых моделей (в более узком плане – искусственных нейронных сетей).

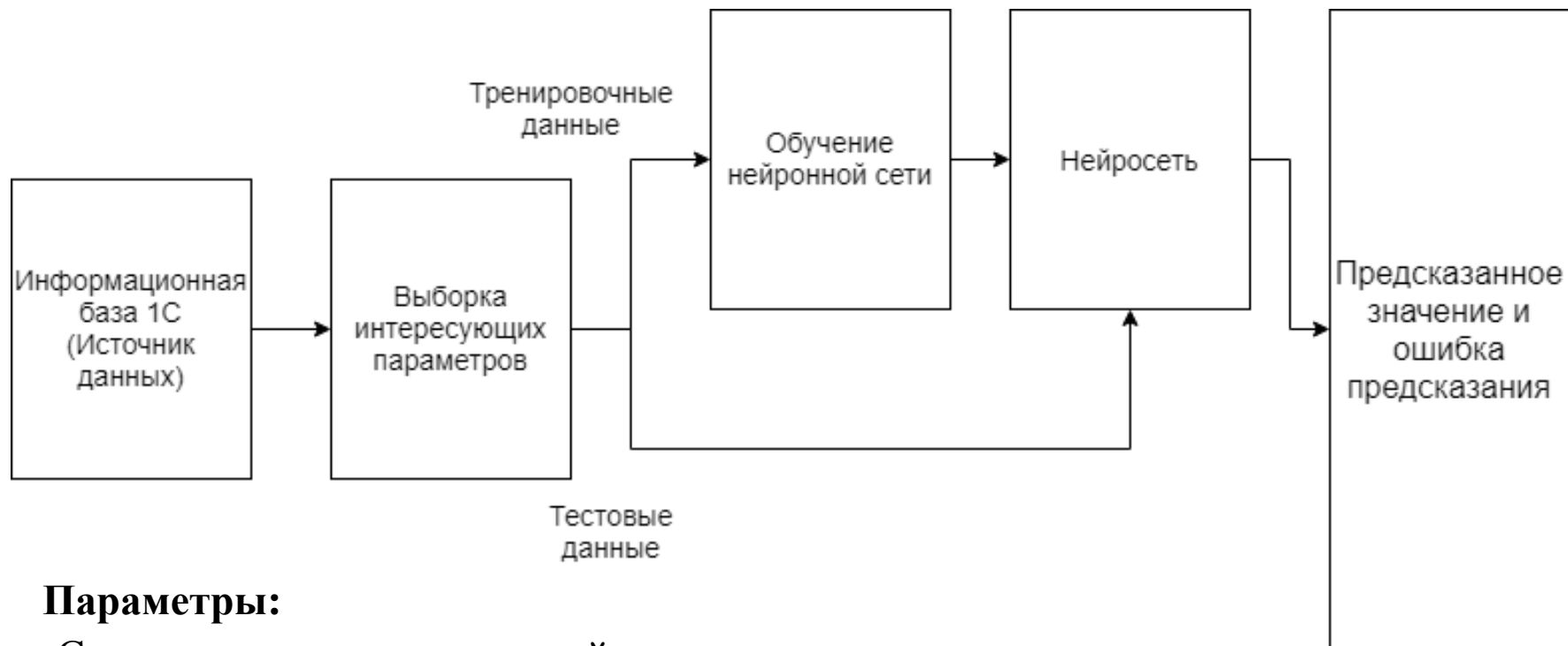


# Особенности функционирования искусственных нейронных сетей (ИНС)

1. Подбор архитектуры сети (количество слоев и нейронов в каждом слое и др.)
2. Обучение ИНС, которое характеризуется целевой функцией, которая направлена на снижение ошибки предсказания и состоит из отдельных «эпох».



# Обработка данных с использованием искусственных нейронных сетей

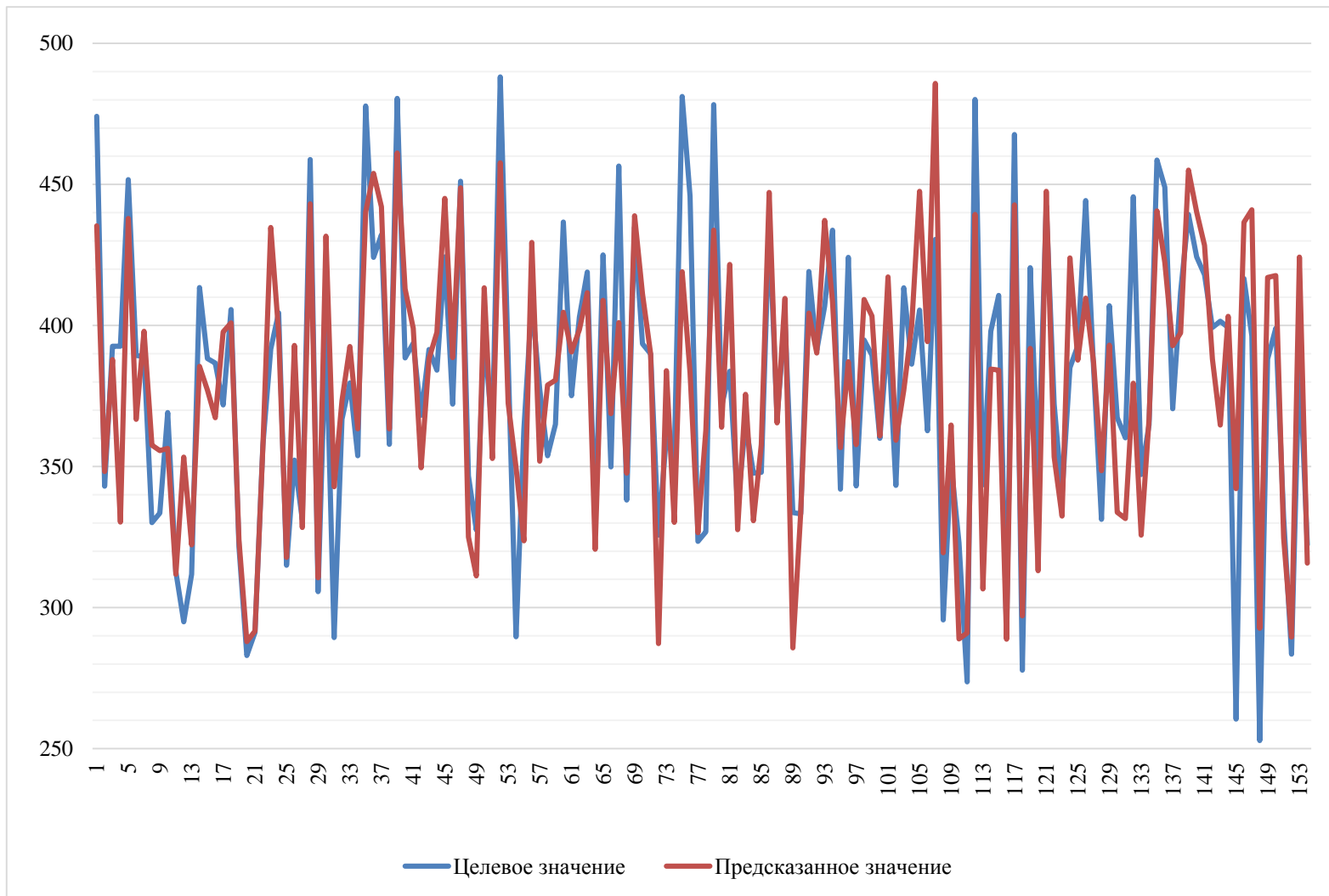


## Параметры:

- Степень помола макулатурной массы в напорном ящике бумагоделательной машины (БДМ);
- рН макулатурной массы в напорном ящике;
- температура массы в напорном ящике;
- концентрация регистровой воды;
- средняя масса 1 м<sup>2</sup> готовой продукции;

В качестве целевого значения для предсказания используется значение показателя «Абсолютное сопротивление продавливанию».

# Результат прогнозирования показателя «Абсолютное сопротивление продавливанию»





## Расчет ошибки предсказания

Ошибка предсказания рассчитана по формуле:

$$O = \frac{|Y - \hat{Y}|}{Y} * 100, \%$$

Где

$O$  – ошибка предсказания, %;

$Y$  – фактическое значение, кПа;

$\hat{Y}$  – предсказанное значение, кПа;

Ошибка, %	Количество значений, шт.	Количество значений, %.	Макс. ошибка предсказания, кПа
Менее 2	33	21,4	± 7,6
2-4	36	23,4	± 15,2
5-7	41	26,6	± 26,5
8-10	22	14,3	± 37,9
10 и более	22	14,3	

## Выводы

1. Использование нейронных сетей позволяет довольно точно (ошибка предсказания не больше 7% для 71,4% предсказанных значений) предсказывать фактические значения отдельных физико-механических показателей.
2. Расширение списка входных технологических параметров для тренировки и тонкая настройка параметров самой нейронной сети способны в дальнейшем увеличить точность предсказаний.
3. Получение предсказанных значений on-line возможно при обеспечении достаточных вычислительных мощностей и автоматизации процесса выгрузки данных на вход нейронной сети посредством современных инструментов для работы с Big Data.
4. На основе анализа структуры нейронной сети предполагается разработка системы рекомендаций для технологического персонала для поддержания или увеличения физико-механических показателей продукции в зависимости от текущих параметров технологического режима.

# Yandex

02/13/2019

Денис Николаевич Жирнов

has successfully completed

Big Data Essentials: HDFS, MapReduce and Spark RDD

an online non-credit course authorized by Yandex and offered through Coursera

COURSE  
CERTIFICATE



*Alexey Dral, Ivan Puzirevskiy, Emel Dral, Evgeniy Ryabenko*

Alexey Dral, Ivan Puzirevskiy,  
Emel Dral, Evgeniy Ryabenko

Verify at [coursera.org/verify/6K8W6Ka](https://coursera.org/verify/6K8W6Ka)  
Coursera has confirmed the identity of this individual and their participation in the course.

# Yandex

04/21/2019

Денис Николаевич Жирнов

has successfully completed

Big Data Analysis: Hive, Spark SQL, DataFrames and GraphFrames

an online non-credit course authorized by Yandex and offered through Coursera

COURSE  
CERTIFICATE



*Natalia Pritykovskaya, Kirilenkov, Pavel Mezentsov, Alexey Dral*

Natalia Pritykovskaya, Kirilenkov,  
Pavel Mezentsov, Alexey Dral

Verify at [coursera.org/verify/WYUVQTSBWF5](https://coursera.org/verify/WYUVQTSBWF5)  
Coursera has confirmed the identity of this individual and their participation in the course.

# Спасибо

# за внимание!!!

# Yandex

06/01/2019

Денис Николаевич Жирнов

has successfully completed

Big Data Applications: Machine Learning at Scale

an online non-credit course authorized by Yandex and offered through Coursera

COURSE  
CERTIFICATE



*Vladimir Lesnichenko, Pavel Mezentsov, Emel Dral, Alexey A. Dral, Ilya Trofimov, Evgeny Frolov*

Vladimir Lesnichenko, Pavel Mezentsov, Emel Dral, Alexey A. Dral, Ilya Trofimov, Evgeny Frolov

Verify at [coursera.org/verify/L4VU6PRM5WRL](https://coursera.org/verify/L4VU6PRM5WRL)  
Coursera has confirmed the identity of this individual and their participation in the course.