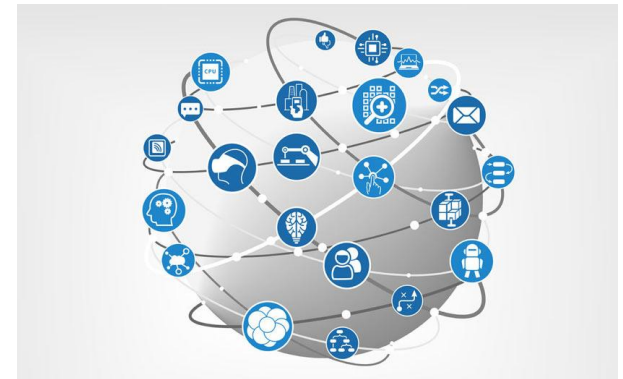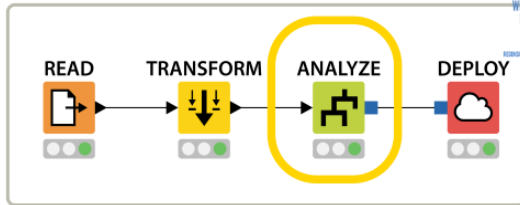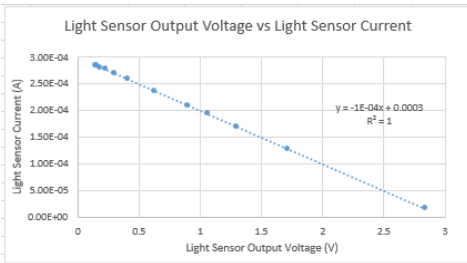# **Predictive Analytics and Machine Learning Basics**

Class 2: Creating a Predictive Analytic Model

July 17, 2018
Don Wilcher

Class 2: Creating a Predictive Analytic Model

Agenda

- The Predictive Analytics Process
- What is Linear Regression
- Exploring Orange
- Lab Project: Predicting the electrical behavior of a light sensor circuit.

Presented by:

DesignNews

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

The Predictive Analytics Process

There are seven steps to the Predictive Analytics Process.

1. **Define Project**
2. **Data Collection**
3. **Data Analysis**
4. **Statistics**
5. **Modeling**
6. **Deployment**
7. **Model Monitoring**

Presented by:

The Predictive Analytics Process...

1. **Project Definition** – Identifying the project outcomes based on the design objectives specified.

2. **Data Collection** – Acquiring and preparing data for the purpose of predicting future trends for processing and monitoring systems.

3. **Data Analysis** – The process of reviewing data results using statistical methods.

The Predictive Analytics Process...

4. Statistics – The validation of assumptions and hypothesis using statistical methods.

5. Modeling – The ability to create predictive models that can accurately forecast future trends.
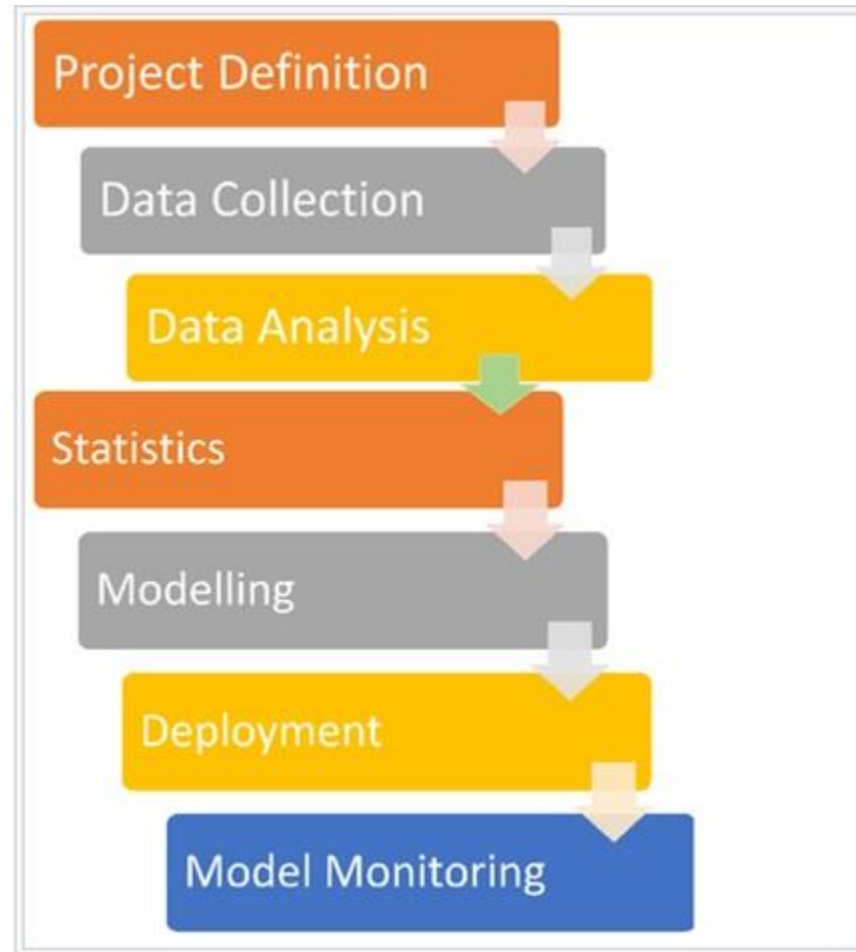
6. Deployment – To send out the predictive model for use in decision making processes, tasks and activities.

7. Model Monitoring – The management and monitoring the predictive performance of the analytical model.

Presented by:

DesignNews

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

The Predictive Analytics Process…

Predictive Analytics (PA) Process Model

Project Definition

Data Collection

Data Analysis

Statistics

Modelling

Deployment

Model Monitoring

What is Linear Regression?

Linear Regression – Identifying the project outcomes based on the design objectives specified.

FAQs:

- The focus is on predicting a value of Y having been given the value of X.

- The regression line is called the *prediction line*.

- The regression line is a straight line that lies closet to all the points in the scatter plot.
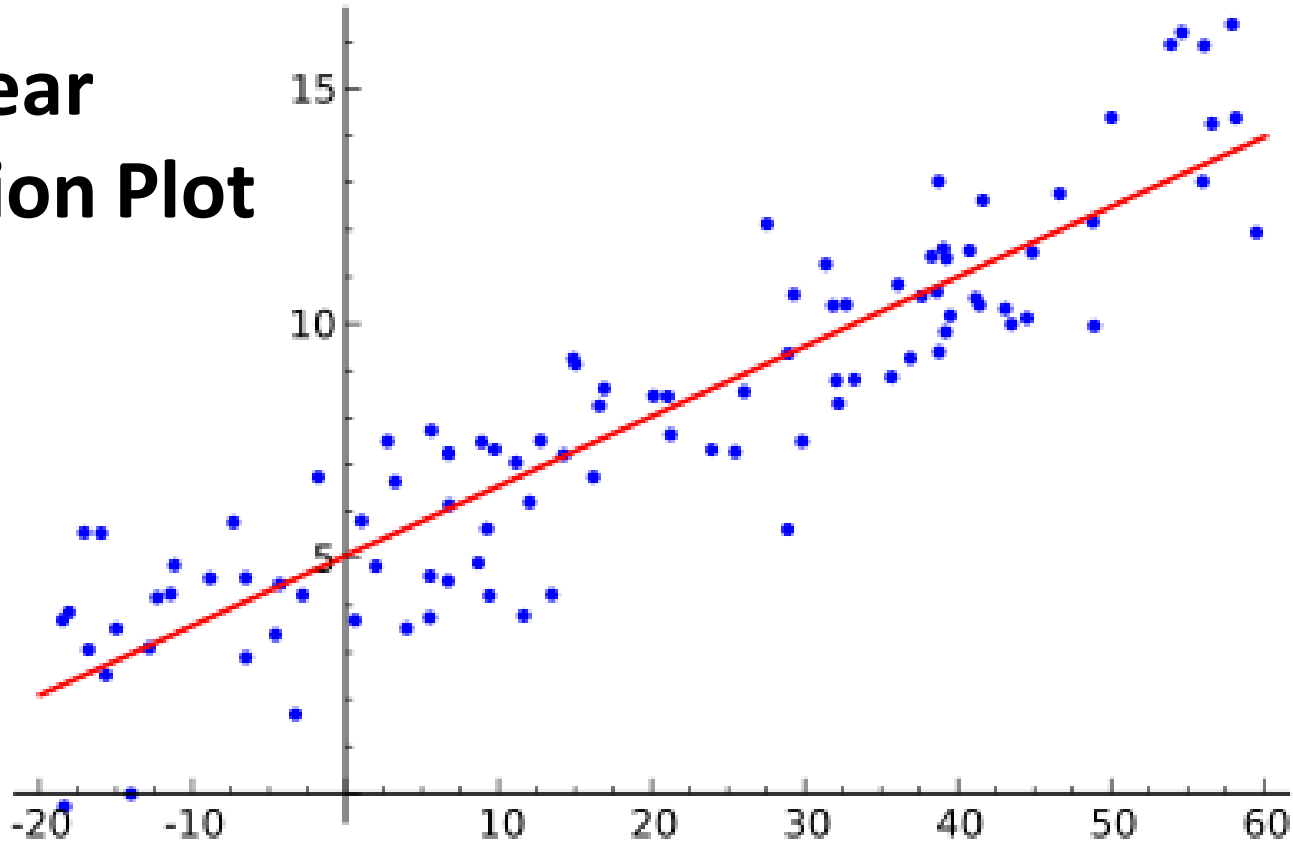
Source:
Sprinthall, R. (1987). Basic statistical analysis, 3rd ed. Englewood Cliffs, NJ: Prentice Hall Inc.

Presented by:

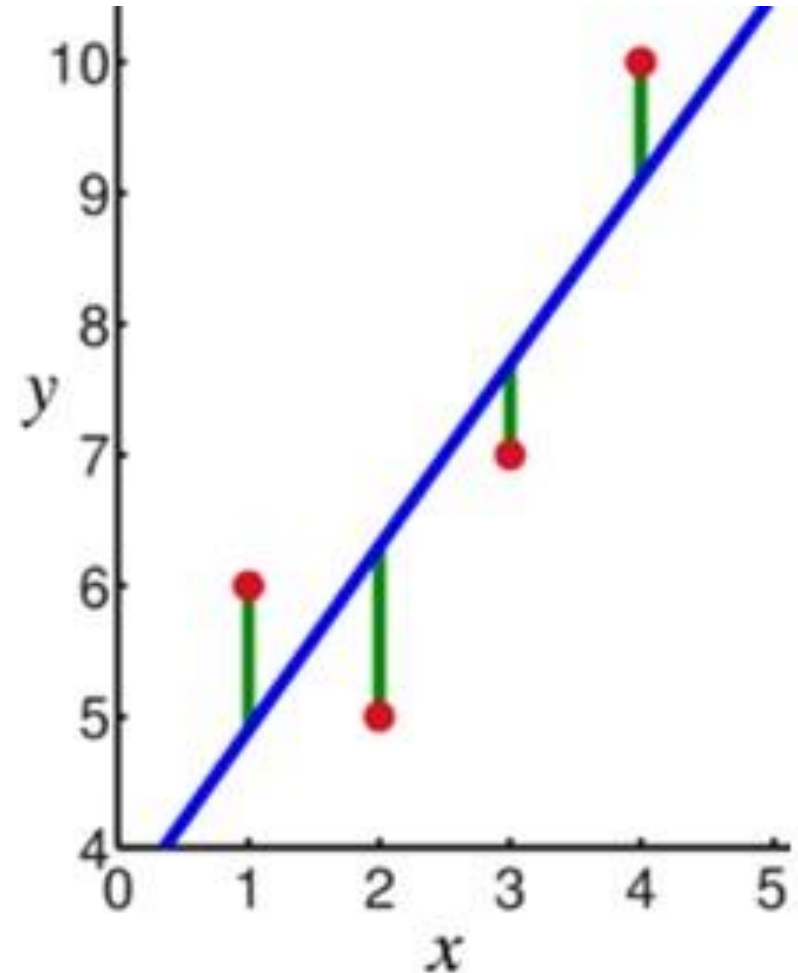# What is Linear Regression?. . .

Linear Regression Plot

Question 1

True or False: There are eight steps for the Predictive Analytics Model process.

Presented by:

What is Linear Regression?. . .

Linear Regression Plot Anatomy

a)Red dots: observation

b)Green line: deviation

c)Blue line: *y* vs *x*

What is Linear Regression?. . .

The Extent of the Scatter Around the Regression

- The closer the points on the scatter plot cluster around the regression line, the higher is the resulting correlation between x and y.

- The closer the points to the regression line, the more accurate is the resulting prediction.

- The higher the correlation, the closer the scatter points cluster around the regression line.
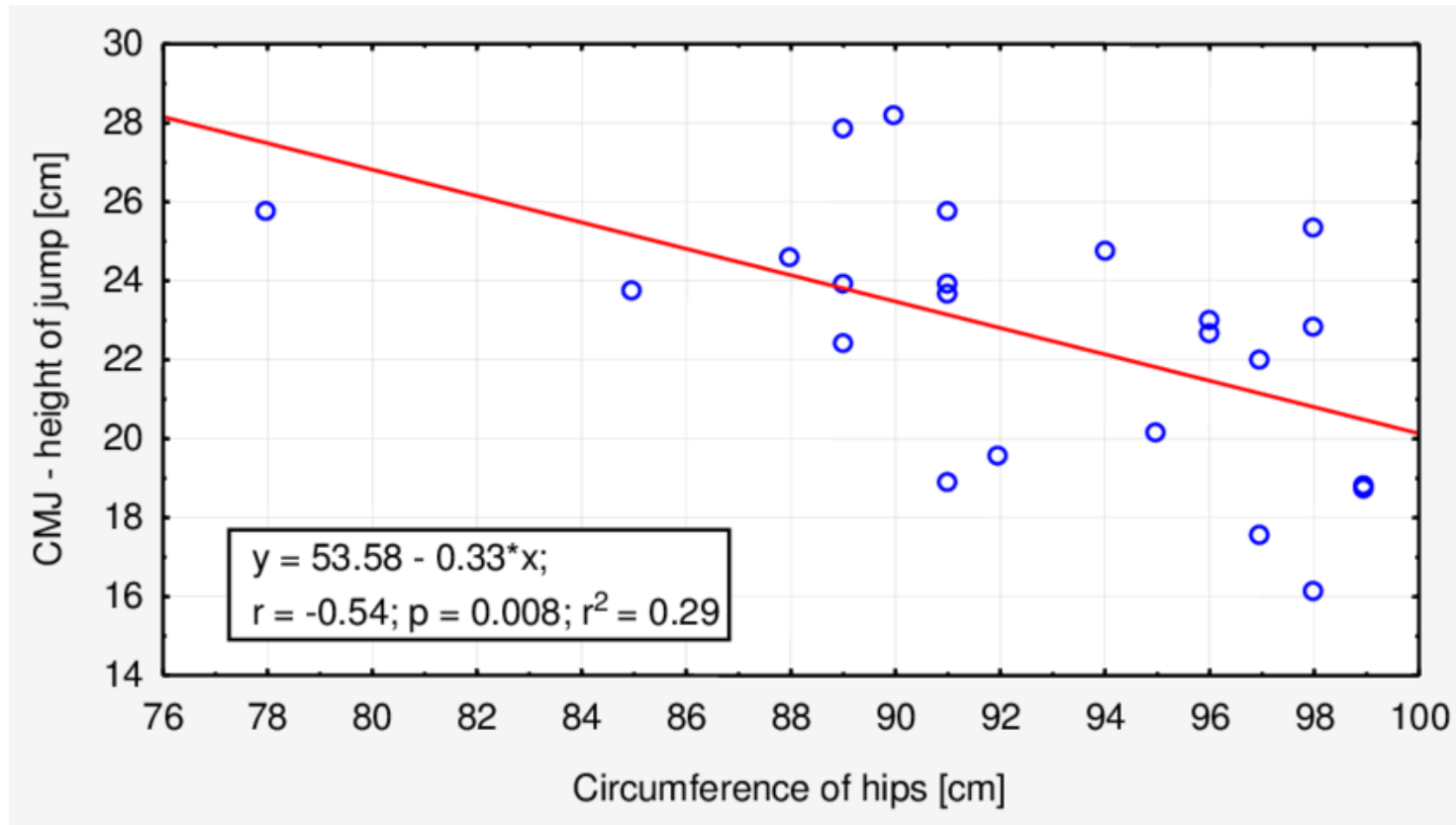
Source:
Sprinthall, R. (1987). Basic statistical analysis, 3rd ed. Englewood Cliffs, NJ: Prentice Hall Inc.

Presented by:

DesignNews

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

What is Linear Regression?. . .

Scatter plot: Circumference of hips [cm] (x-axis, 76 to 100) vs CMJ - height of jump [cm] (y-axis, 14 to 30).

$$y = 53.58 - 0.33*x;$$
$$r = -0.54; \quad p = 0.008; \quad r^2 = 0.29$$

Source:

https://www.researchgate.net/publication/309526456_Application_of_the_index_of_balance
-stiffness_for_evaluation_of_the_process_of_maintaining_body_balance/figures?lo=1
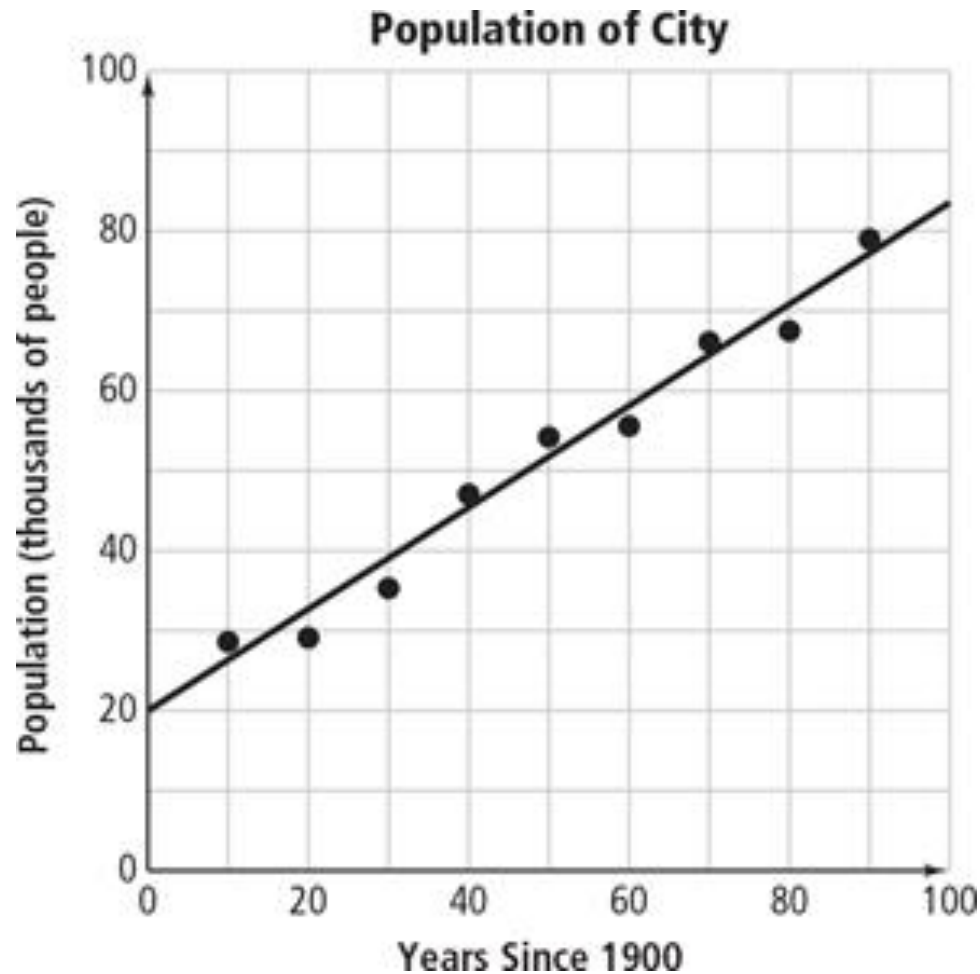
Presented by:

What is Linear Regression?. . .

Population of City

Presented by:

13

Question 2

True or False: Linear Regression is a complex Machine Learning Model for Predictive Analytics.

Presented by:

What is Linear Regression?. . .

Linear Regression based on this simple equation.

$$y = mx + b$$

output — slope — input — y-intercept

Note:

and   are coefficients for the linear equation

Presented by:

Exploring Orange

orange   Features   Screenshots   Download   Docs   Blog   Training   Donate

Data Mining Fruitful and Fun

Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

Download Orange

Source:
https://orange.biolab.si/

DesignNews

16

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS
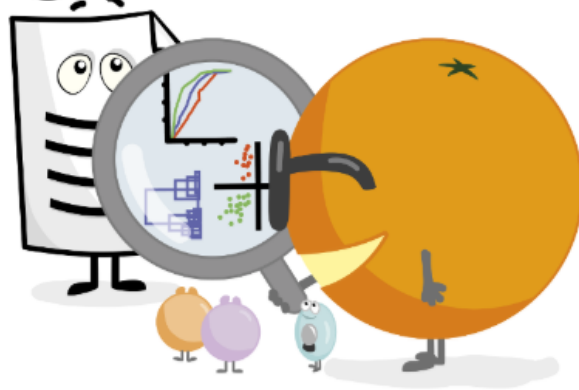
Exploring Orange. . .

Interactive Data Visualization

Perform simple data analysis with clever data visualization. Explore statistical distributions, box plots and scatter plots, or dive deeper with decision trees, hierarchical clustering, heatmaps, MDS and linear projections. Even your multidimensional data can become sensible in 2D, especially with clever attribute ranking and selections.
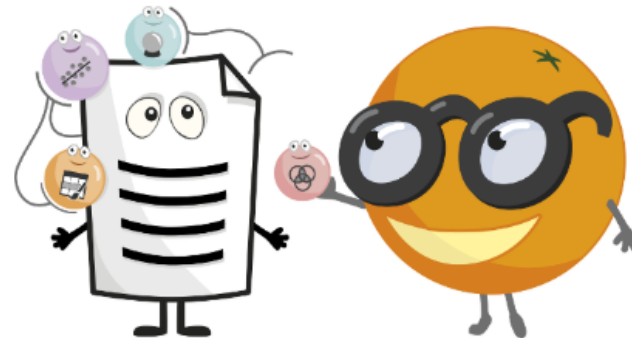
⊕ Learn More

Visual Programming

Interactive data exploration for rapid qualitative analysis with clean visualizations. Graphic user interface allows you to focus on exploratory data analysis instead of coding, while clever defaults make fast prototyping of a data analysis workflow extremely easy. Place widgets on the canvas, connect them, load your datasets and harvest the insight!

⊕ Learn More   ⊙ Watch Video

Source:
https://orange.biolab.si/

Presented by:

Exploring Orange. . .

Presented by:

Exploring Orange. . .

orange

Documentation

Visual Programming

Getting started
YouTube tutorials
Loading your data
Widget catalog

Development

Widget development
Example addon

Python Library

Tutorial
Reference
Orange 2.7 documentation

Source:
https://orange.biolab.si/docs/

Presented by:

19

Question 3

Orange is an open source machine learning and _____ visualization tool for novice and expert developers.

Exploring Orange. . .

Presented by:

21

Exploring Orange. . .

Analyzing a Cricket's Chirp Rate vs Temperature: Workflow

Presented by:

DesignNews

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS
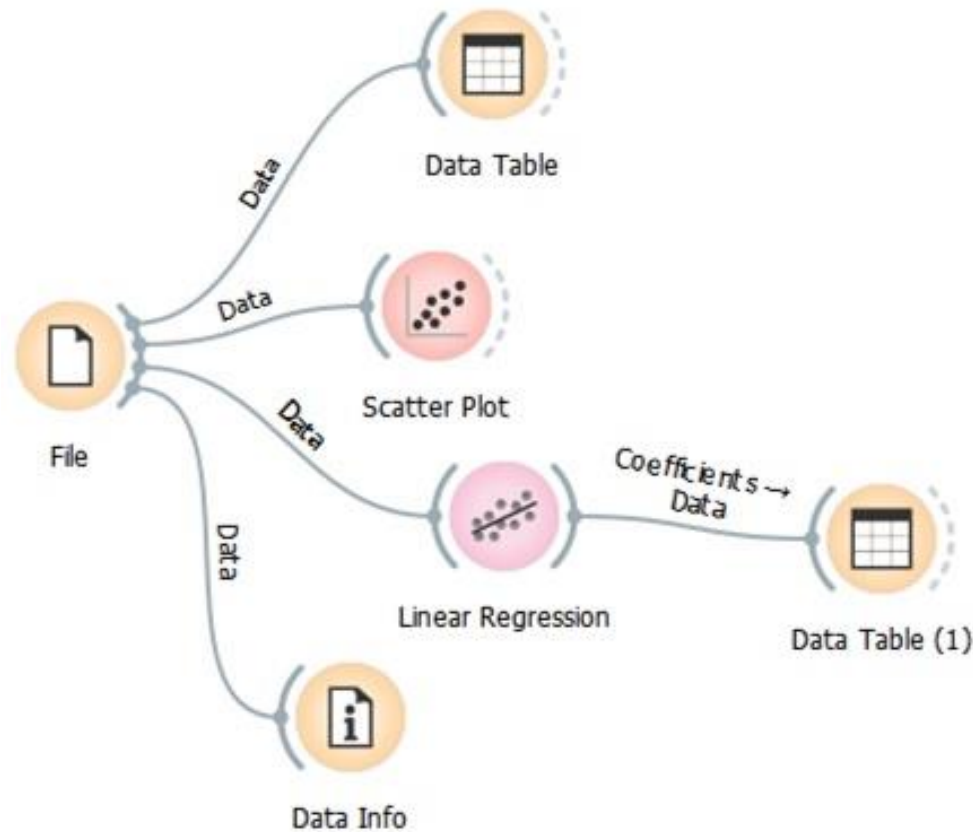
Exploring Orange. . .

Cricket's Chirp Rate vs Temperature Data:

CSV file

```
Chirp Rate,Temperature
20,88.6
16,71.6
19.8,93.3
18.4,84.3
17.1,80.6
15.5,75.2
14.7,69.7
17.1,82
15.4,69.4
16.3,83.3
15,79.6
17.2,82.6
16,80.6
17,83.5
14.4,76.3
```

Using slide 23: In configuring the input file, what feature is used as the target for the machine learning model in Orange?

Presented by:

Exploring Orange. . .

Data Table window:

Info
- 15 instances (no missing values)
- 1 feature (no missing values)
- Continuous target variable (no missing values)
- No meta attributes

Variables
- ☑ Show variable labels (if present)
- ☐ Visualize continuous values
- ☑ Color by instance classes

Selection
- ☑ Select full rows

	Temperature	Chirp Rate
1	88.600	20.000
2	71.600	16.000
3	93.300	19.800
4	84.300	18.400
5	80.600	17.100
6	75.200	15.500
7	69.700	14.700
8	82.000	17.100
9	69.400	15.400
10	83.300	16.300
11	79.600	15.000
12	82.600	17.200
13	80.600	16.000
14	83.500	17.000
15	76.300	14.400

Restore Original Order

Report

☑ Send Automatically

Presented by:

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

Exploring Orange. . .

Presented by:

Exploring Orange. . .

Data Table (1)

Info

2 instances (no missing values)
1 feature (no missing values)
No target variable.
1 meta attribute (no missing values)

	name	coef
1	intercept	24.9660144
2	Chirp Rate	3.3057614

Variables

☑ Show variable labels (if present)
☐ Visualize continuous values
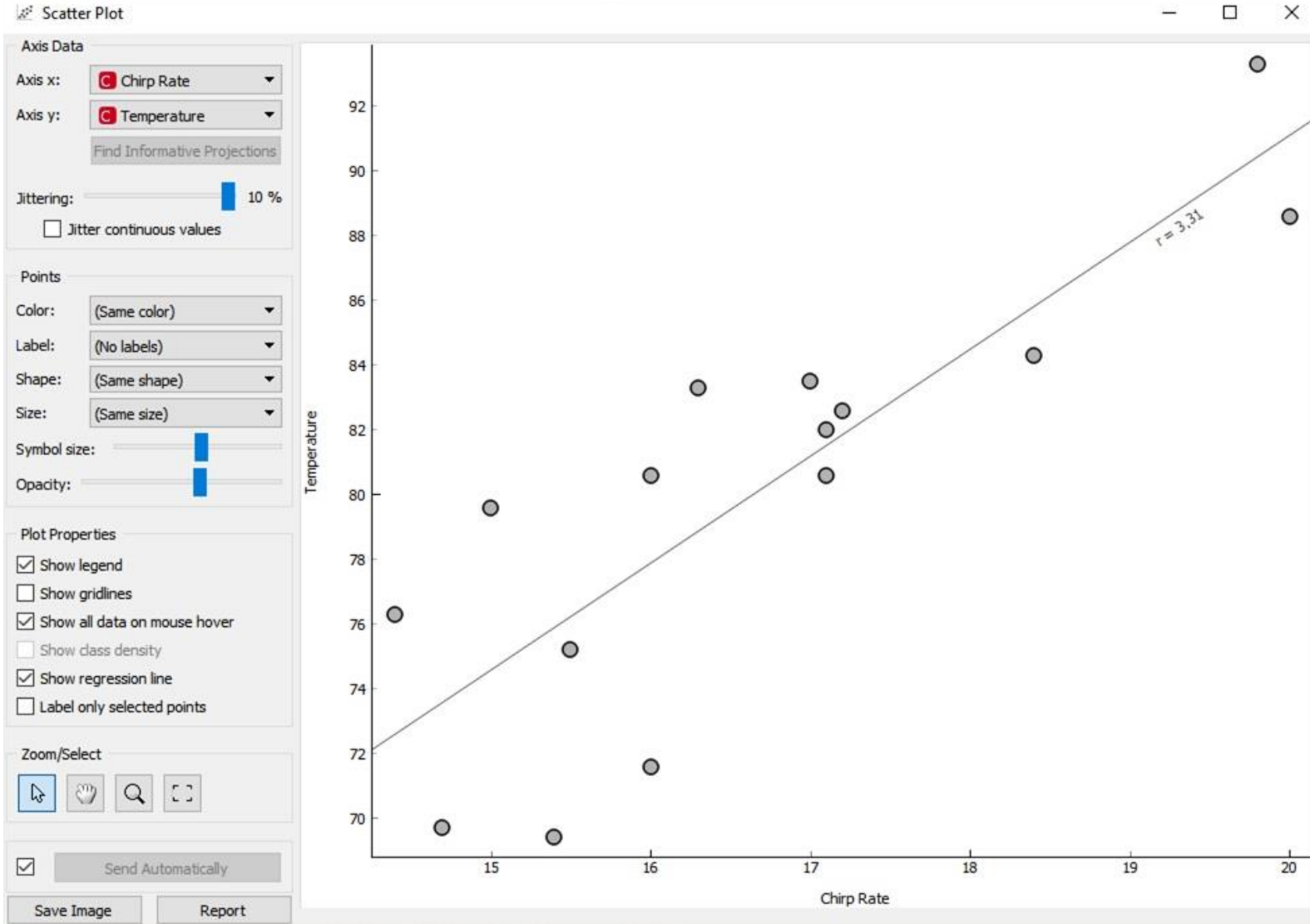☑ Color by instance classes

Selection

☑ Select full rows

Restore Original Order

Report

☑ Send Automatically

Linear Regression equation.

$$y = 3.3057614x + 24.9660144$$

Presented by:

DesignNews

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

Exploring Orange. . .

Predicting a new temperature with a future chirp rate example.

Given:

$$x = 21 Hz$$

$$y = 3.3057614x + 24.9660144$$

$$y = 3.3057614(21) + 24.9660144$$

$$y = 94.3869744 \,°F$$

Presented by:

Question 5

Using slide 28: The numbers in the linear regression expression are the _____ of the equation.

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit

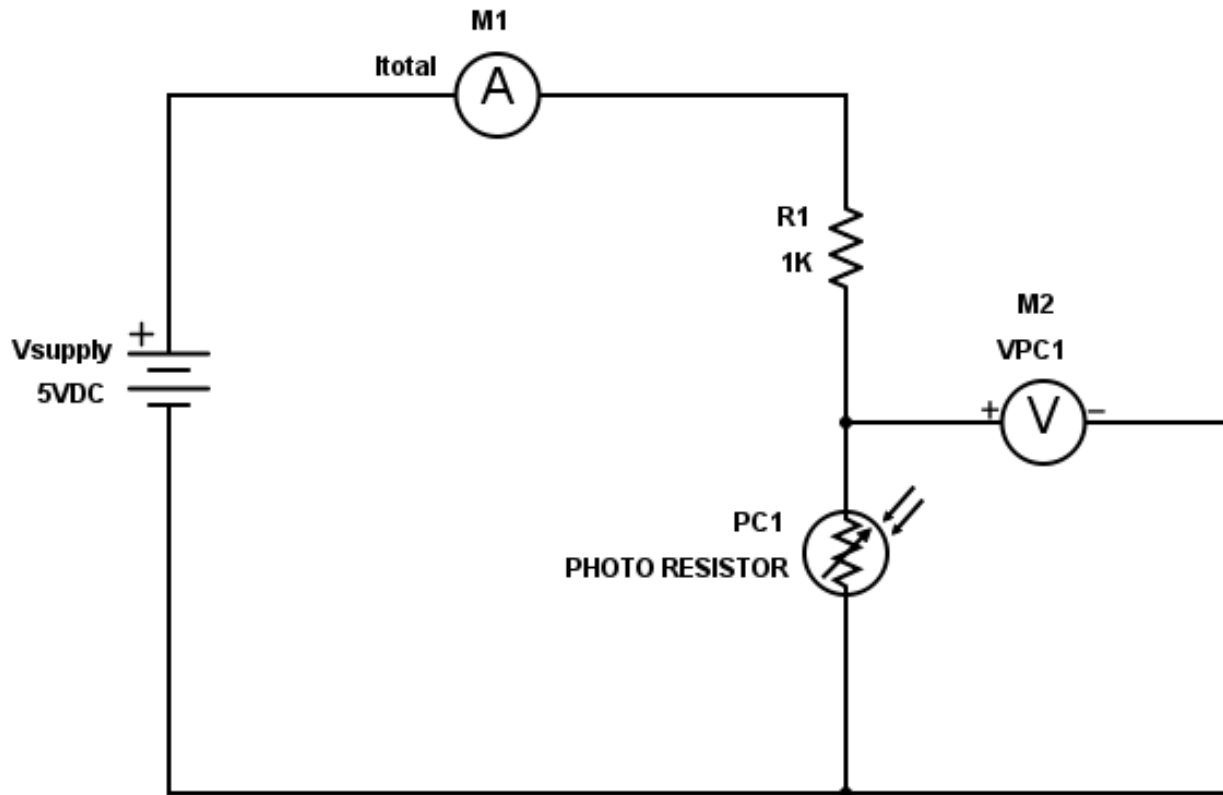**Lab Project Objectives:**

a. Learn how to build an electronic sensor in TinkerCad Circuits.

b. Learn how to create Common Separated Values file.

c. Learn how to build a Predictive Model using Orange.

Presented by:

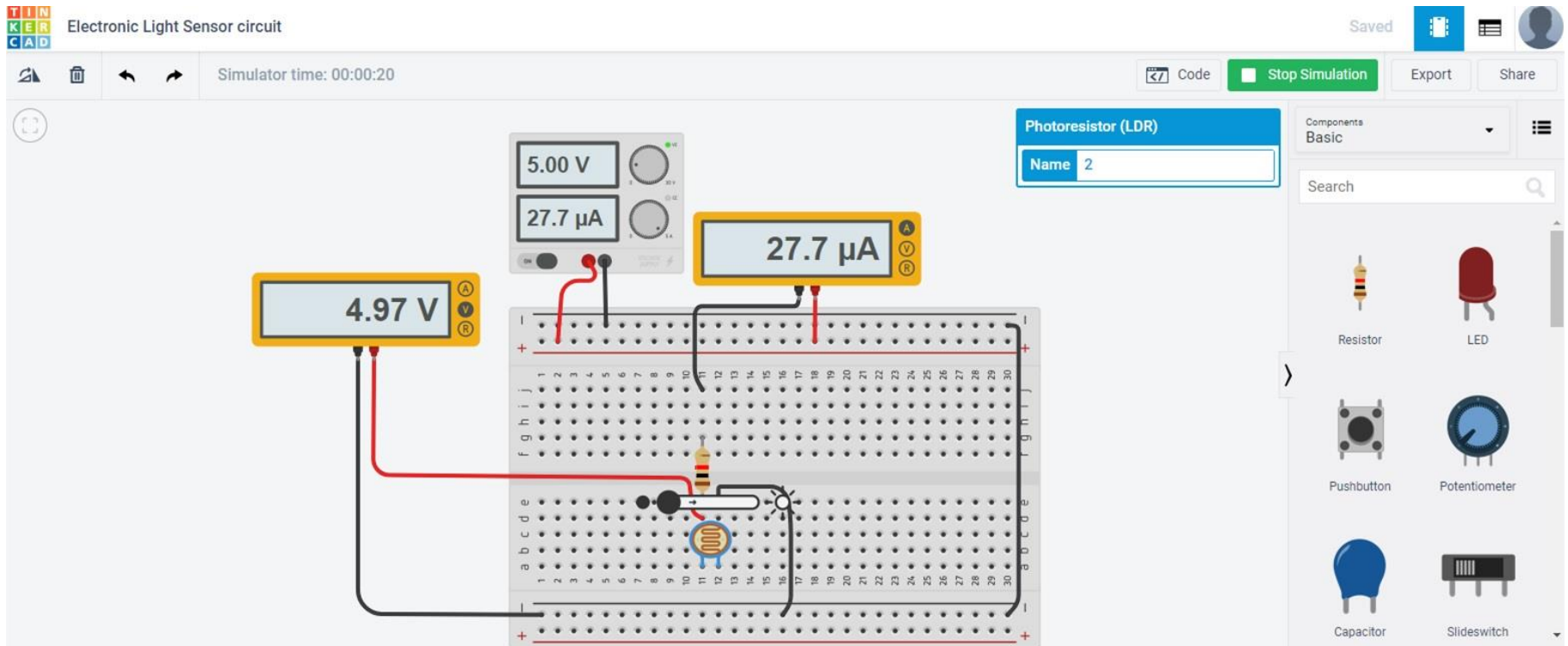# Predicting the Electrical Behavior of a Light Sensor Circuit. . .

Electronic Light Sensor Circuit.

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

TinkerCad Circuits: Electronic Light Sensor Circuit

Circuit Link:
https://www.tinkercad.com/things/aUIgFjfHDw5-electronic-light-sensor-circuit/editel

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

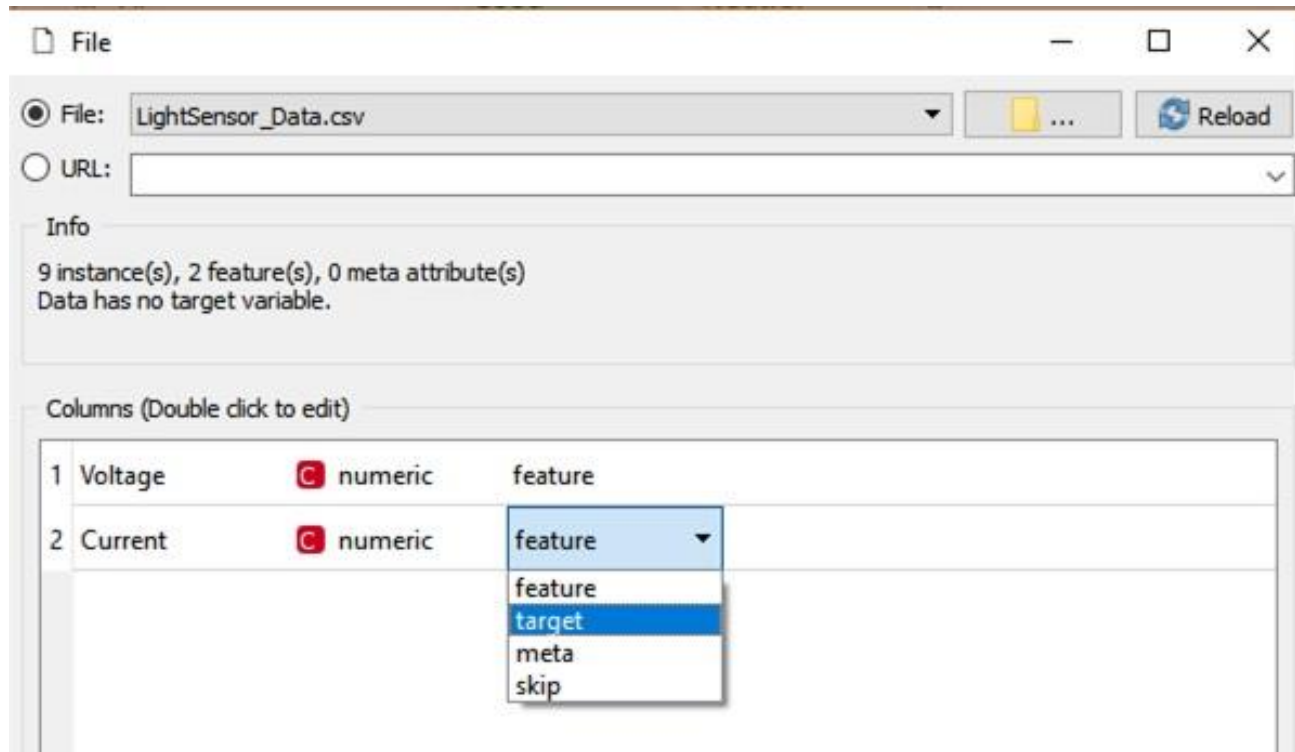## Electronic Light Sensor Circuit Data(9 datapoints)

```
Voltage,Current
4.97,2.77E-05
4.42,5.78E-04
4.43,7.74E-04
4.06,9.45E-04
3.90,1.10E-03
3.64,1.36E-03
3.53,1.47E-03
3.32,1.68E-03
2.92,2.21E-03
```

CSV file created in notepad

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

Establishing an Electronic Light Sensor Circuit target variable.

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

	Current	Voltage
1	0.000028	4.97
2	0.000578	4.42
3	0.000774	4.43
4	0.000945	4.06
5	0.001100	3.90
6	0.001360	3.64
7	0.001470	3.53
8	0.001680	3.32
9	0.002210	2.92

Data Table

Info
9 instances (no missing values)
1 feature (no missing values)
Continuous target variable (no missing values)
No meta attributes

Variables
☑ Show variable labels (if present)
☐ Visualize numeric values
☑ Color by instance classes
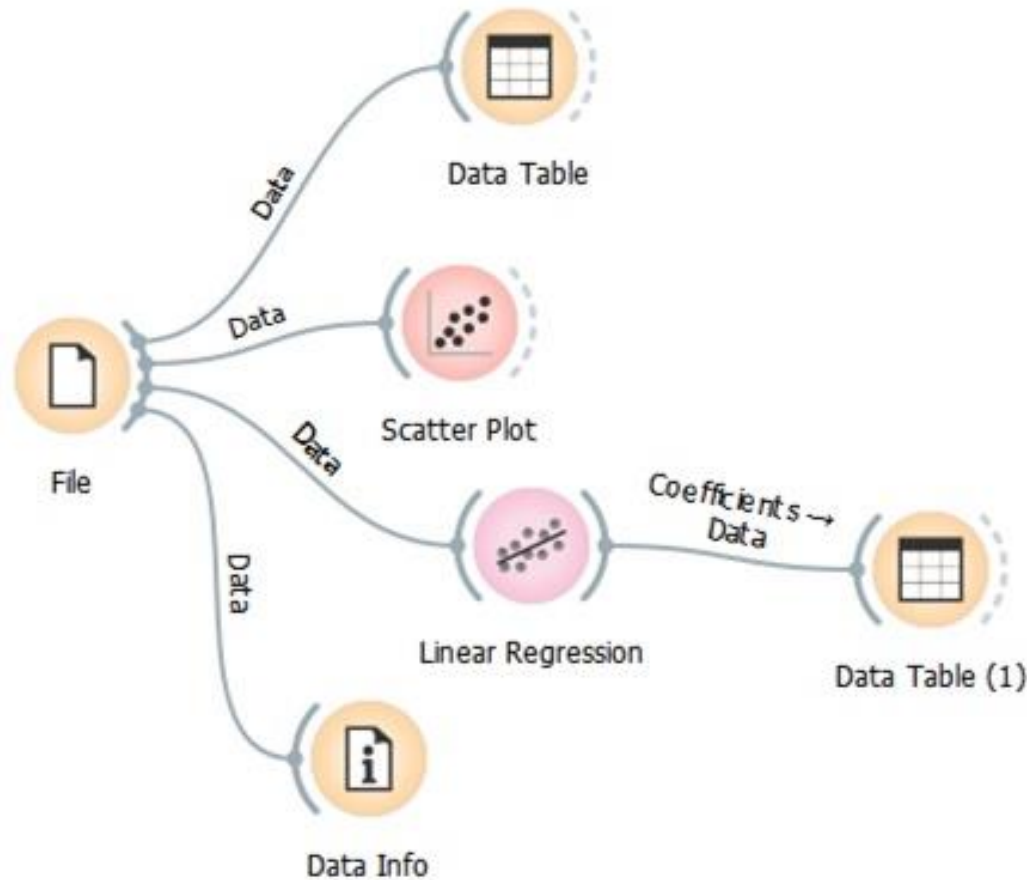
Selection
☑ Select full rows

Restore Original Order

☑ Send Automatically

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

Electronic Light Sensor Circuit Predictive Model Workflow
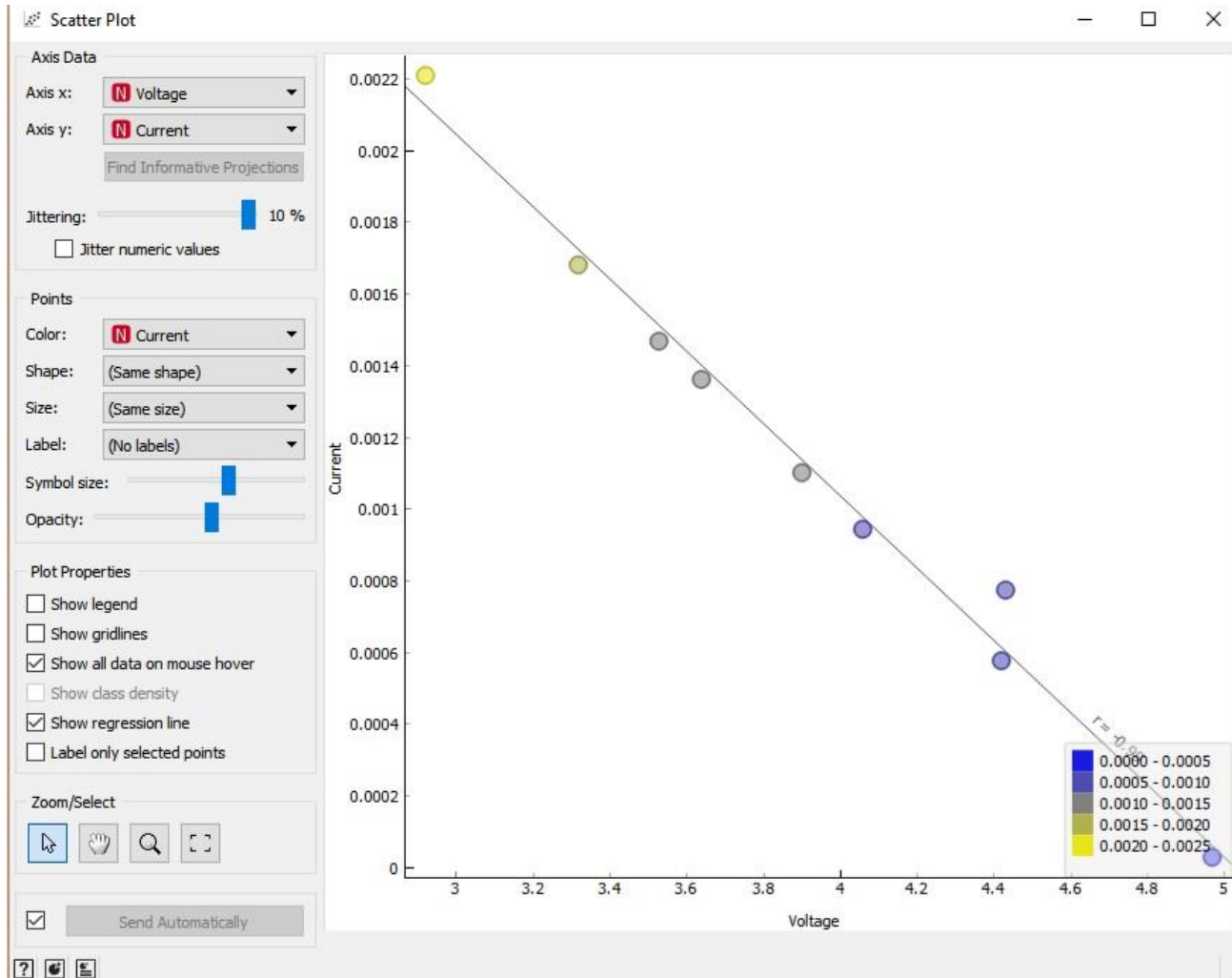
Using slide 34: In configuring the input file, what feature is used as the target for the machine learning model in Orange?

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

	name	coef
1	intercept	0.0050679
2	Voltage	-0.0010079

Linear Regression equation.

$$y = -0.0010079x + 0.0050679$$

Data Table (1)

Info

2 instances (no missing values)
1 feature (no missing values)
No target variable.
1 meta attribute (no missing values)

Variables

☑ Show variable labels (if present)
☐ Visualize continuous values
☑ Color by instance classes

Selection

☑ Select full rows

Restore Original Order

Report

☑ Send Automatically

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

Predicting a new current value with a electronic sensor output voltage example.

Given:

$$x = 1.68V$$

$$y = -0.0010079x + 0.0050679$$
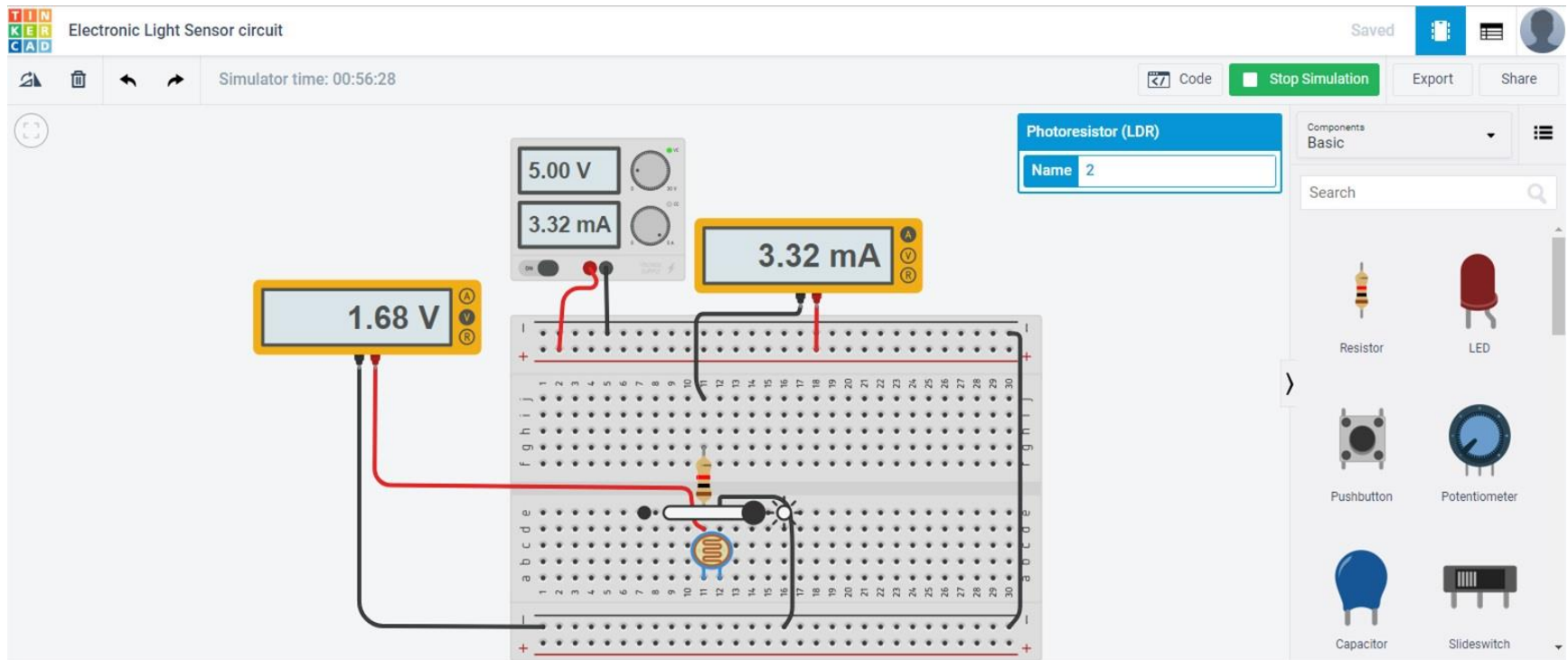
$$y = -0.0010079(1.68) + 0.0050679$$

$$y = 0.003374628A = 3.374628mA$$

Presented by:

Predicting the Electrical Behavior of a Light Sensor Circuit. . .

Predictive output (Current) measured on the electronic light sensor model.

Presented by:

Data Collection with a Raspberry Pi. . .

Running (execute) the write file application in Python:

Console Results

Presented by:

What is the name of the online circuit modeling tool for constructing the Electronic Light Sensor Circuit?

Presented by:

Data Collection with a Raspberry Pi. . .

Contents of person.csv.txt

/home/pi/person.csv.txt - pi@192.168.7.76 - Editor - WinSCP

```
Name of Person,Age
Peter,22
Jasmine,21
Sam,24
Don,32
```

Presented by:

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

DesignNews