

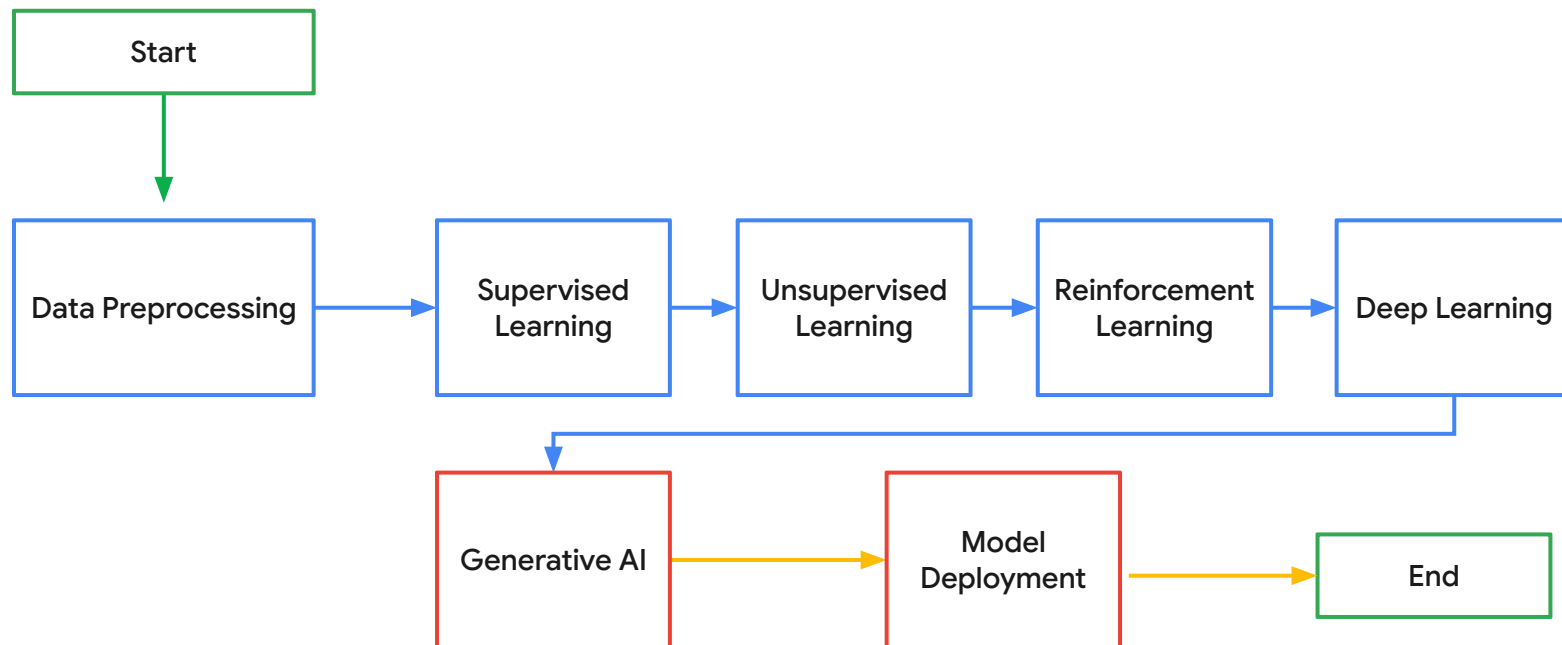


# Introduction to Supervised Learning and Data Preprocessing



Ahmad Daffa Abiyyu  
Academy Team, GDGoC  
UNAIR

# The Curriculum



# What are we going to learn today:

1. Data Preprocessing
2. Basic Supervised Models  
(part 1)



# Data Preprocessing

```
lookup.KeyValue  
f.constant(['em  
=tf.constant([G  
.lookup.StaticV  
_buckets=5)
```

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
  )  
),  
),  
Text(  
  'Ka  
  sty  
  c  
),  
),  
),
```

# Importing and Viewing Data Information



## Google Developer Groups

**“Our baby step, yet  
our most important  
step”**

```
er(
l(32),

/*1*/
child: Column(
    crossAxisAlignment: CrossAxisAlignment
    children: [
        /*2*/
        Container(
            padding: const EdgeInsets
            child: const Text(
                'Oeschinen Lake Campg
                style: TextStyle(
                    fontWeight: FontWei
            ),
        ),
    ],
)
```

# EXAMPLE

# Handling Value Format in a Column





## Google Developer Groups

**“Format in column  
could sometime  
fooled us”**

```
er(
l(32),

/*1*/
child: Column(
    crossAxisAlignment: CrossAxisAlignment
    children: [
        /*2*/
        Container(
            padding: const EdgeInsets
            child: const Text(
                'Oeschinen Lake Campg
                style: TextStyle(
                    fontWeight: FontWei
            ),
        ),
    ],
)
```

# EXAMPLE

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
    )  
  ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
    ),  
    ),
```

# Handling Missing Value in a Column



Google Developer Groups

```
er(
l(32),

/*1*/
child: Column(
  crossAxisAlignment: CrossAxisAlignment
  children: [
    /*2*/
    Container(
      padding: const EdgeInsets
      child: const Text(
        'Oeschinen Lake Campg
        style: TextStyle(
          fontWeight: FontWei
        ),
      ),
    ),
  ],
)
```

# EXAMPLE

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
    )  
  ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
    ),  
    ),
```

# Data Visualization



**“The most aesthetic aspect of data preprocessing”**

```
er(
l(32),

/*1*/
child: Column(
    crossAxisAlignment: CrossAxisAlignment
children: [
    /*2*/
    Container(
        padding: const EdgeInsets
        child: const Text(
            'Oeschinen Lake Campg
            style: TextStyle(
                fontWeight: FontWei
        ),
    ),
)
```

# EXAMPLE



```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
    )  
  ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
    ),  
  ),  
),
```

# Target Column



**“The thing that we want to predict”**

```

er(
l(32),

/*1*/
child: Column(
  crossAxisAlignment: CrossAxisAlignment
  children: [
    /*2*/
    Container(
      padding: const EdgeInsets
      child: const Text(
        'Oeschinen Lake Campg
        style: TextStyle(
          fontWeight: FontWei
      ),
    ),
  ),
)

```

# EXAMPLE

[illegible]

The image shows a presentation slide with a decorative green border featuring geometric patterns like circles and dots. The central white area contains the word "Heatmap" in a large, bold, black font. In the bottom right corner of the white area is the Google Developer Groups logo, which consists of a colorful icon followed by the text "Google Developer Groups". On the far left edge, there is a vertical strip of black background containing white text that appears to be a code snippet, partially cut off. The text includes "ild: Column(", "crossAxisAlignment: CrossAxisAlignment", "children: [", "/\*2\*/", "Conta", "pad", "chi", "s", ")", "(", "Text(", "'Ka", "sty", "c", ")", "(", ")", "1".

The image shows a presentation slide with a decorative green border featuring geometric patterns like circles and dots. The central white area contains the word "Heatmap" in a large, bold, black font. In the bottom right corner of the white area is the Google Developer Groups logo, which consists of a colorful icon followed by the text "Google Developer Groups". On the far left edge, there is a vertical strip of black background containing white text that appears to be a code snippet, partially cut off. The text includes "ild: Column(", "crossAxisAlignment: CrossAxisAlignment", "children: [", "/\*2\*/", "Conta", "pad", "chi", "s", ")", "(", "Text(", "'Ka", "sty", "c", ")", "(", ")", "1".



# “Column have a twin?? ”

```

er(
l(32),

/*1*/
child: Column(
  crossAxisAlignment: CrossAxisAlignment
  children: [
    /*2*/
    Container(
      padding: const EdgeInsets
      child: const Text(
        'Oeschinen Lake Campg
        style: TextStyle(
          fontWeight: FontWei
      ),
    ),
  ),
)

```

# EXAMPLE

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
      pad  
      chi  
        '  
        s  
      )  
    ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
  ),  
  ),  
),
```

# Encoder



# “Turning object into a number”


```
er(
    ll(32),

/*1*/
child: Column(
    crossAxisAlignment: CrossAxisAlignment
    children: [
        /*2*/
        Container(
            padding: const EdgeInsets
            child: const Text(
                'Oeschinen Lake Campg
                style: TextStyle(
                    fontWeight: FontWei
            ),
        ),
    ),
),
)
```




# EXAMPLE


# Scaler

 Google Developer Groups

# Scaler

 Google Developer Groups

# Scaler

 Google Developer Groups



# “Turning number into a smaller”

```
er(
l(32),

/*1*/
child: Column(
  crossAxisAlignment: CrossAxisAlignment
  children: [
    /*2*/
    Container(
      padding: const EdgeInsets
      child: const Text(
        'Oeschinen Lake Campg
        style: TextStyle(
          fontWeight: FontWei
      ),
    ),
  )
)
```

# EXAMPLE

# Supervised Learning: Part 1

```
lookup.KeyValue  
f.constant(['em  
=tf.constant([G  
lookup.StaticV  
_buckets=5)
```

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
    )  
  ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
    ),  
    ),
```

# Linear Regression



**“The most well known regression technique”**

```
er(
l(32),

/*1*/
child: Column(
  crossAxisAlignment: CrossAxisAlignment
  children: [
    /*2*/
    Container(
      padding: const EdgeInsets
      child: const Text(
        'Oeschinen Lake Campg
        style: TextStyle(
          fontWeight: FontWeig
      ),
    ),
  ],
)
```

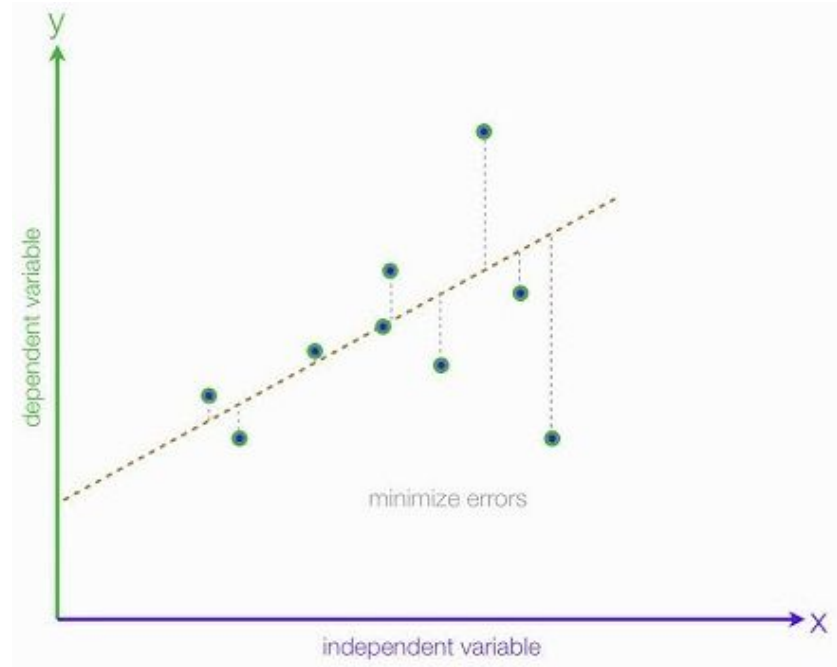
Equation:

$$y = a_1 + a_0 * x + e$$

Where:

$$a_0 = \frac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x^2} - \bar{x}^2}$$

$$a_1 = \bar{y} - a_0 * \bar{x}$$





# EXAMPLE

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
    )  
  ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
    ),  
    ),
```

# Logistic Regression



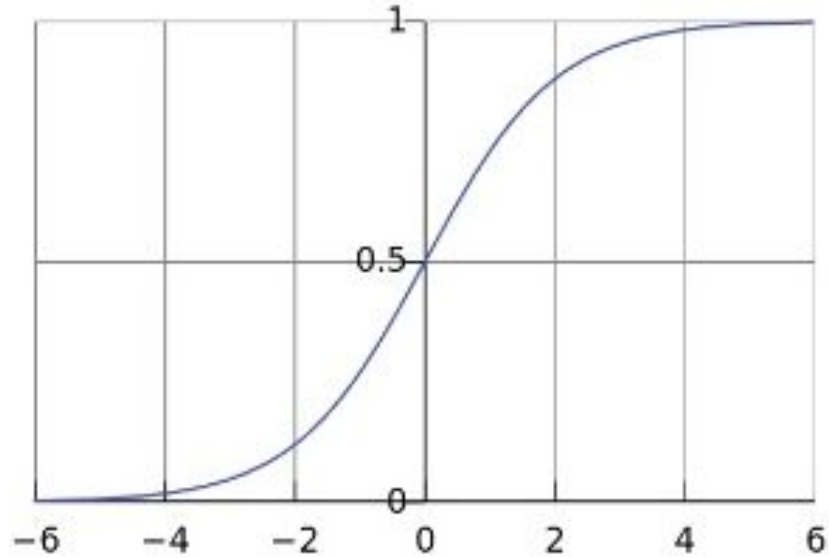
Google Developer Groups

# “The cousin of linear regression”

```
/*1*/  
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Container(  
      padding: const EdgeInsets.  
      child: const Text(  
        'Oeschinen Lake Campg  
        style: TextStyle(  
          fontWeight: FontWeig  
      ),  
    ),  
  ],  
),
```

Equation:

$$S(x) = \frac{1}{1 + e^{-x}}$$



# EXAMPLE

```
child: Column(  
  crossAxisAlignment: CrossAxisAlignment.  
  children: [  
    /*2*/  
    Conta  
    pad  
    chi  
    '  
    s  
    )  
  ),  
  ),  
  Text(  
    'Ka  
    sty  
    c  
    ),  
    ),
```

# Decision Tree



# “The most coolest tree”

```
er(
    ll(32),
    // *1*/
    child: Column(
      crossAxisAlignment: CrossAxisAlignment
    children: [
      // *2*/
      Container(
        padding: const EdgeInsets
      child: const Text(
        'Oeschinen Lake Campg
        style: TextStyle(
          fontWeight: FontWei
        ),
      ),
    ),
  ),
),
```

Equation:

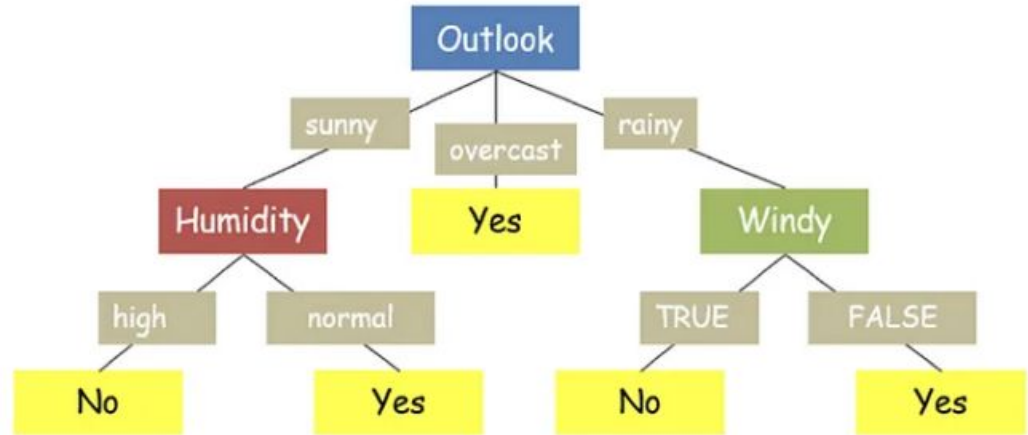
$$IG_{classification} = E(d) - \sum \frac{|s|}{|d|} E(s)$$

$$IG_{regression} = Var(d) - \sum \frac{|s|}{|d|} Var(s)$$

Depends on:

$$Entropy = - \sum_{i=0}^n p_i \log_2 p_i$$

$$Index\ gini = 1 - \sum_i p_i^2$$





# EXAMPLE

Now it's time for...

# Hands on project