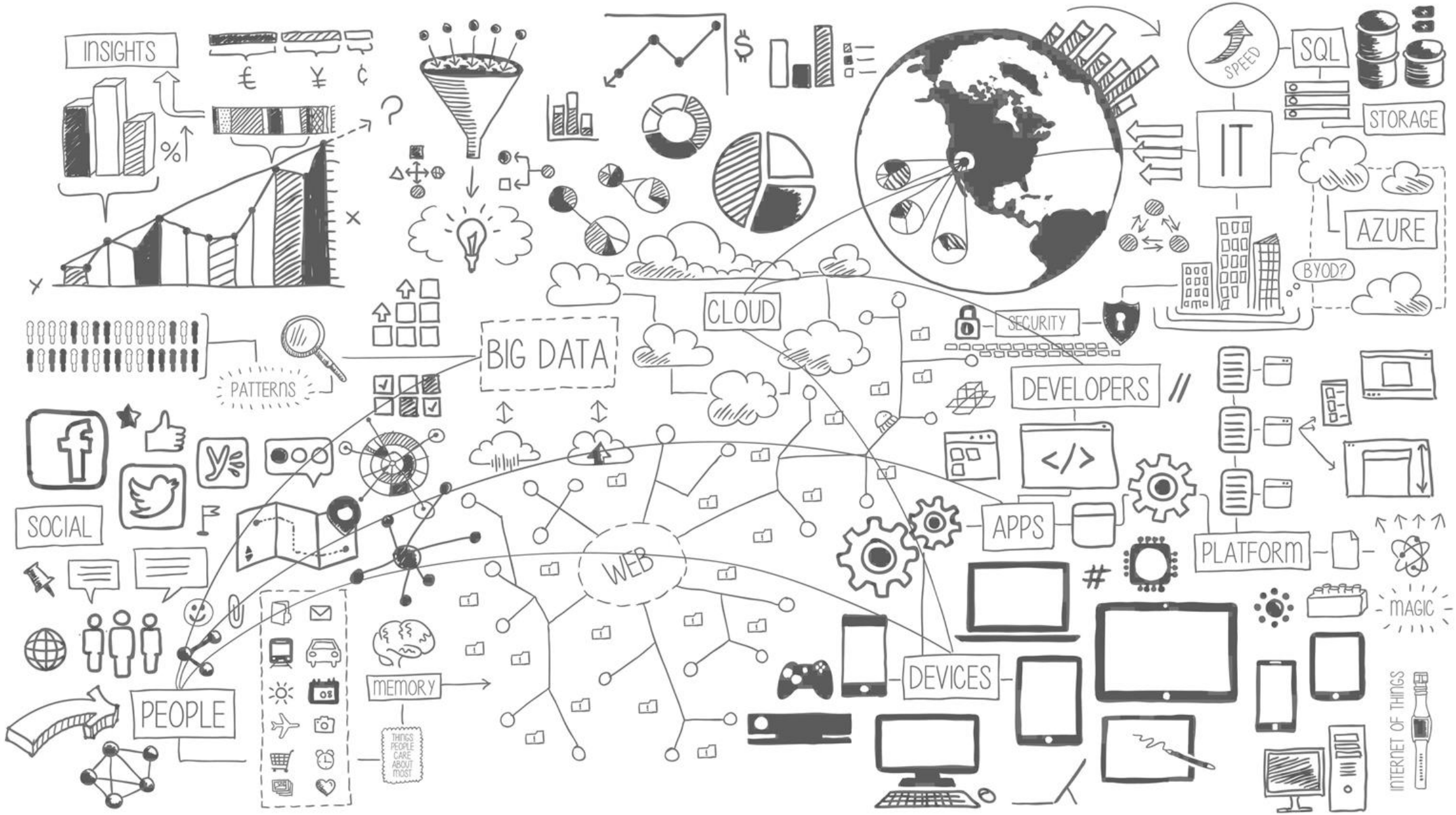
Machine Learning demystified:

do you ask the right questions?

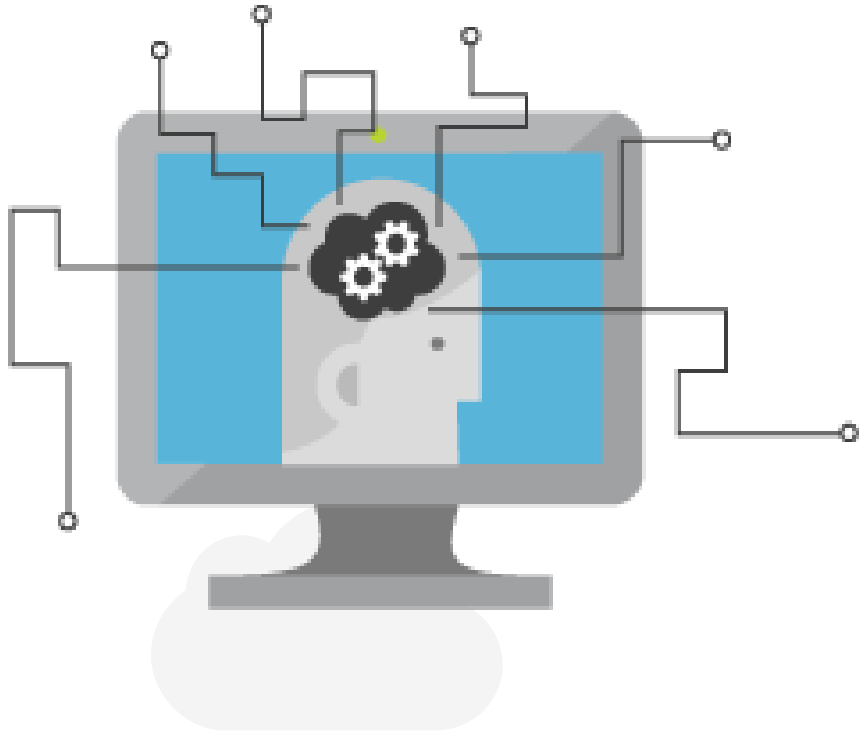
Bianca Furtuna

Technical Evangelist

@Fur_Bi

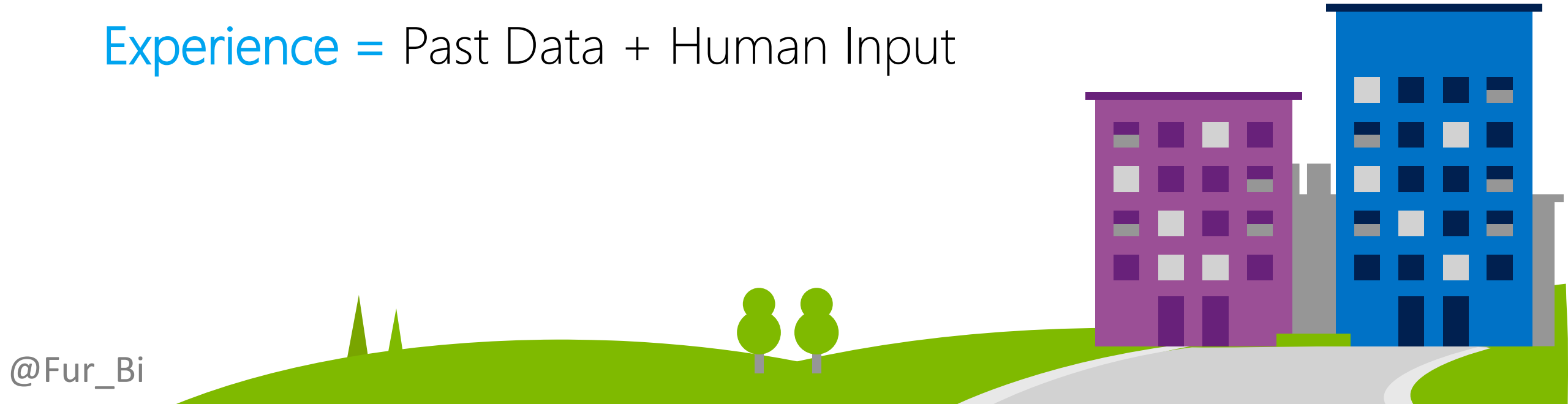

INSIGHTS
SPEED
SQL
STORAGE
IT
AZURE
BYOD?
CLOUD
SECURITY
BIG DATA
PATTERNS
DEVELOPERS
SOCIAL
APPS
PLATFORM
WEB
MAGIC
PEOPLE
MEMORY
THINGS PEOPLE CARE ABOUT MOST
DEVICES
INTERNET OF THINGS

What is Machine Learning?

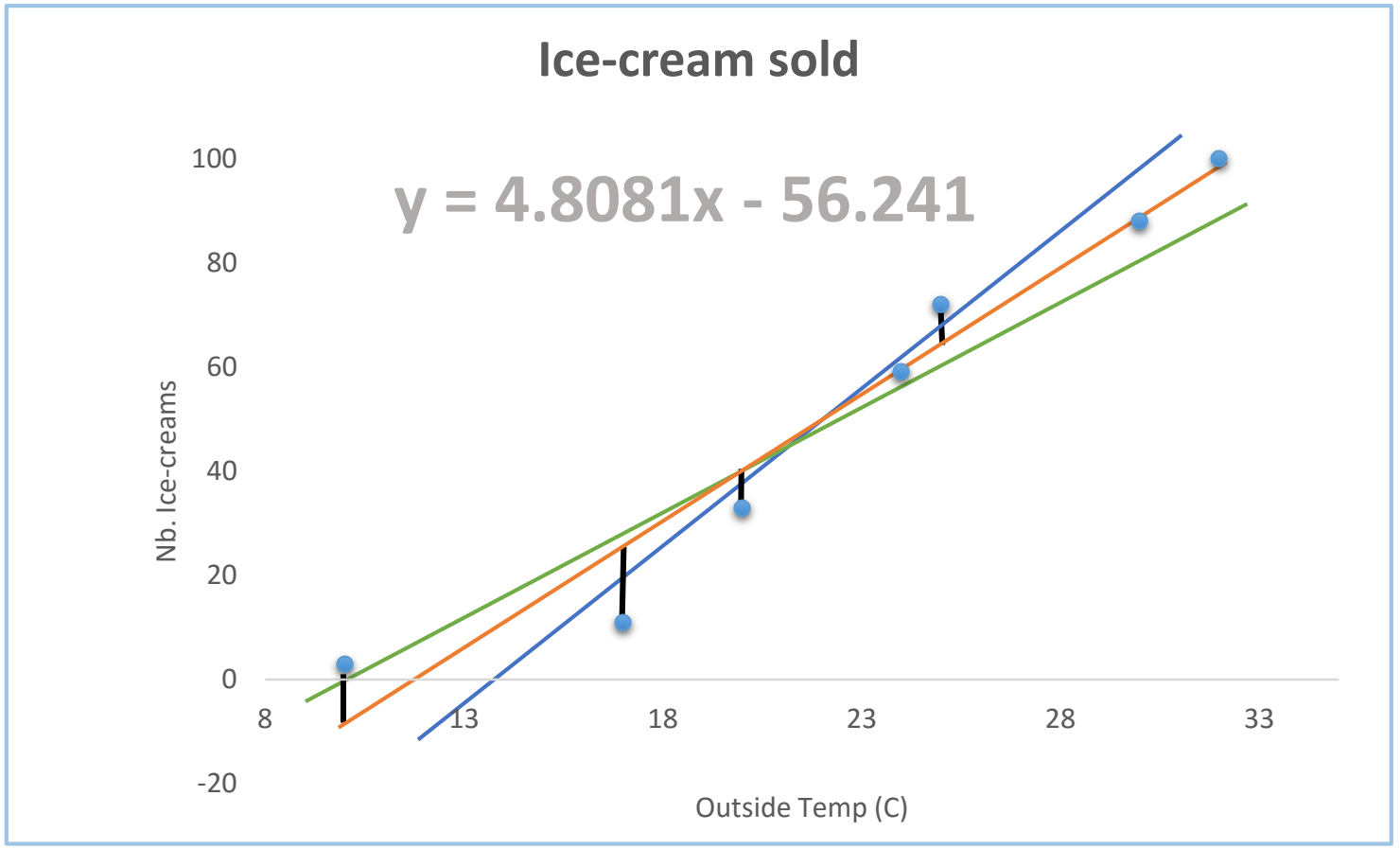


Computing Systems that become smarter with Experience

Experience = Past Data + Human Input

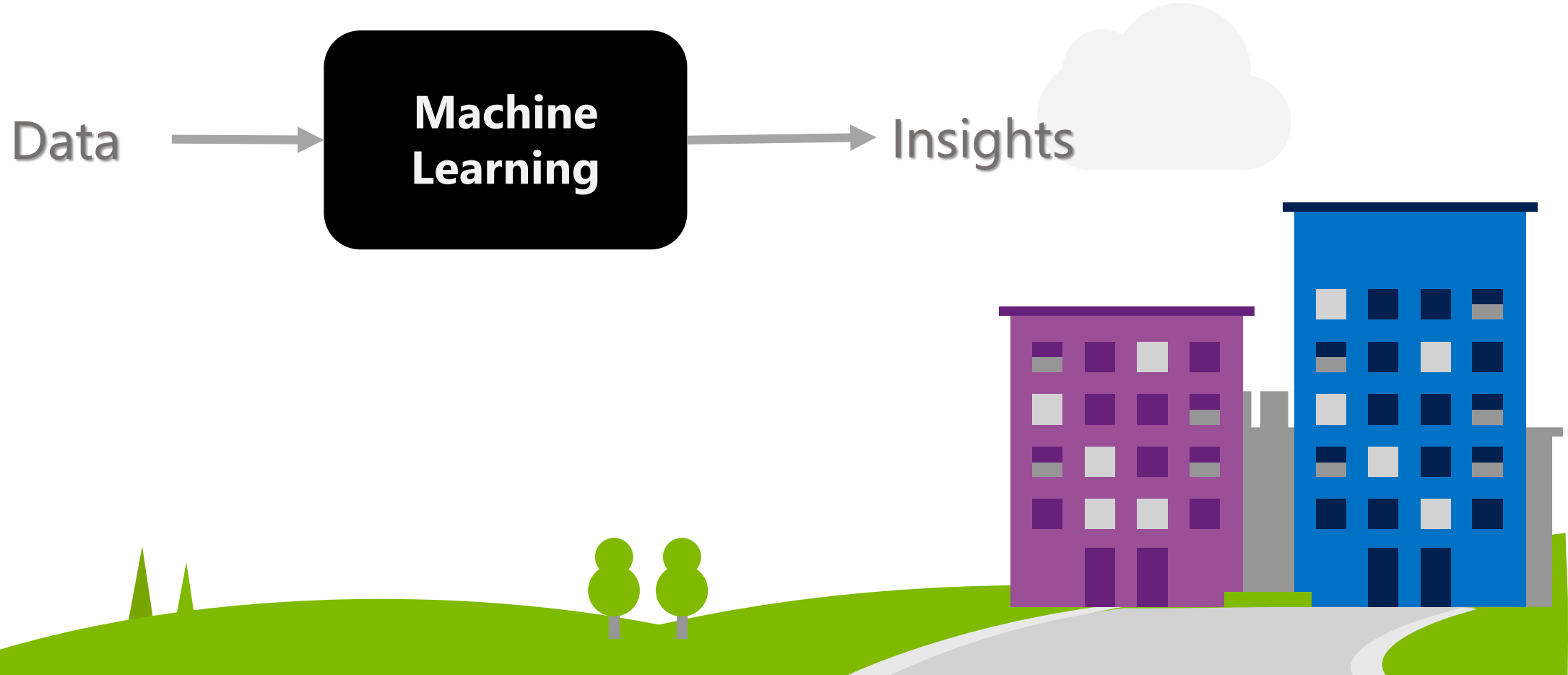


Learning?



$$Y = aX - b$$

How can Machine Learning bring value?

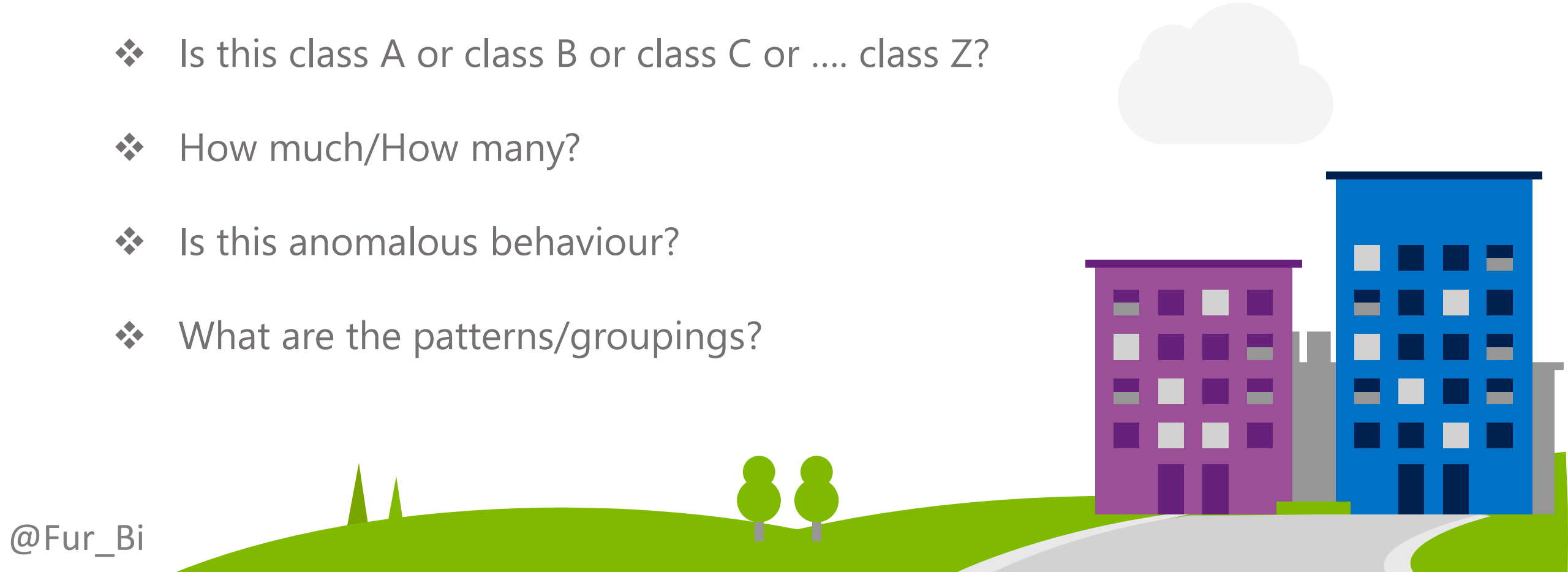


Questions?

Use maths to answer questions

What type of questions?

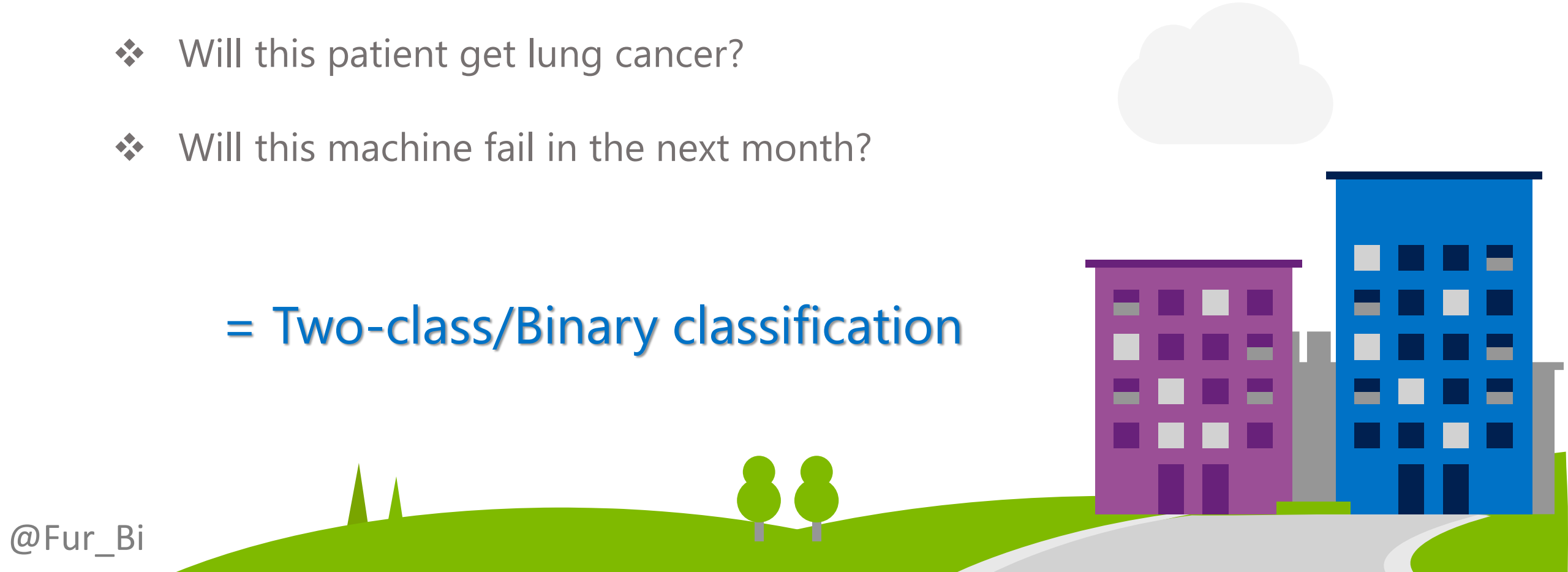
- Is this class A or class B?
- Is this class A or class B or class C or …. class Z?
- How much/How many?
- Is this anomalous behaviour?
- What are the patterns/groupings?



Is this class A or class B?

- Is there a face in this image or not?
- Will this patient get lung cancer?
- Will this machine fail in the next month?

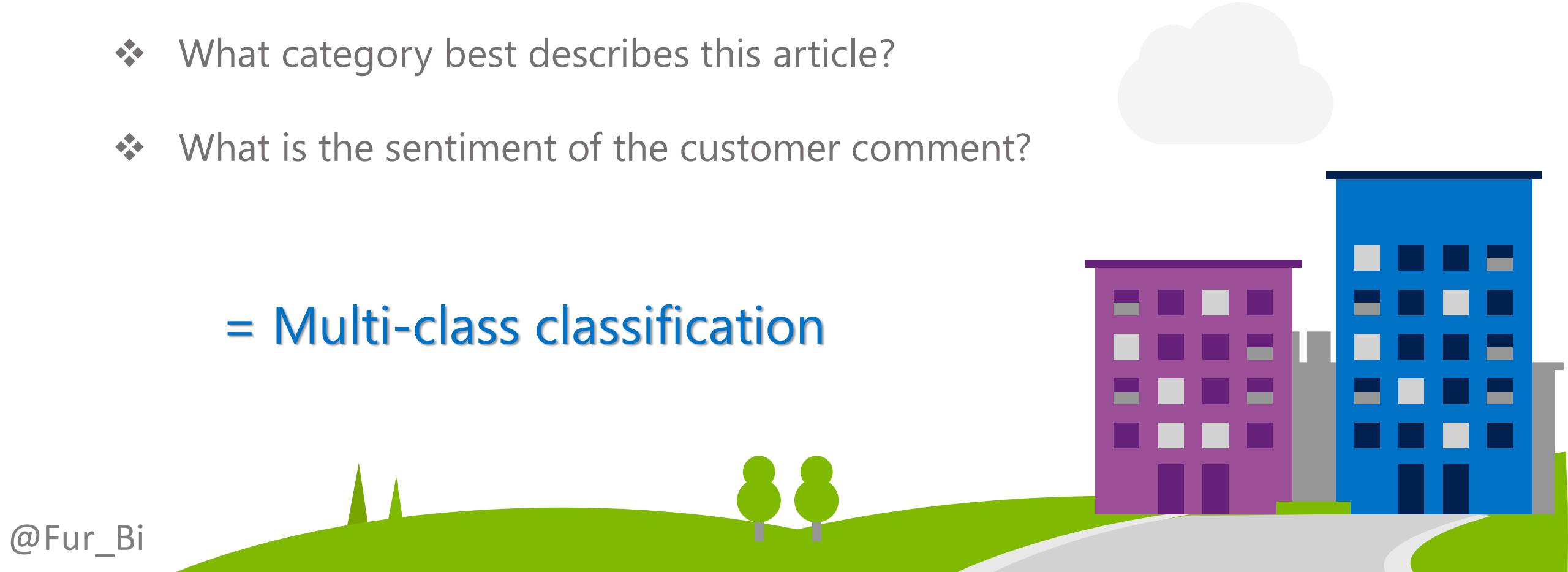
= Two-class/Binary classification



Is this class A or class B or class C or …. class Z?

- What object is in this image?
- What category best describes this article?
- What is the sentiment of the customer comment?

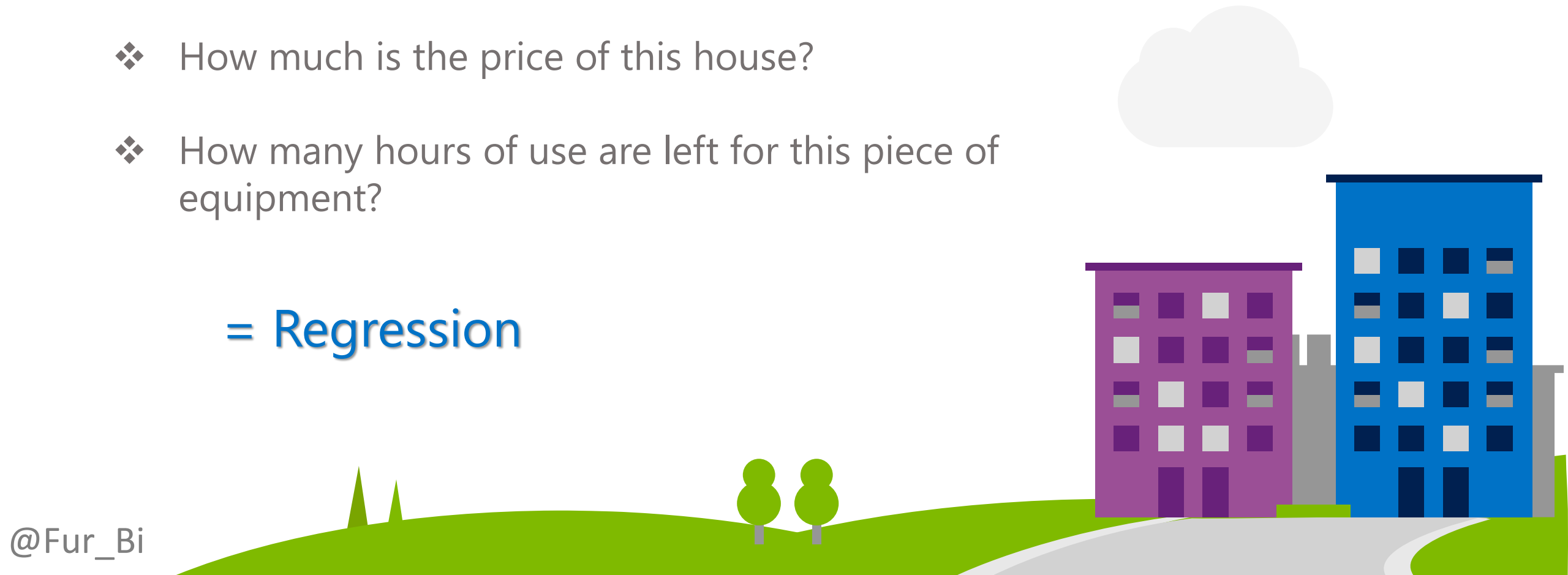
= Multi-class classification



How much/How many?

- How much ice-cream will be sold?
- How much is the price of this house?
- How many hours of use are left for this piece of equipment?

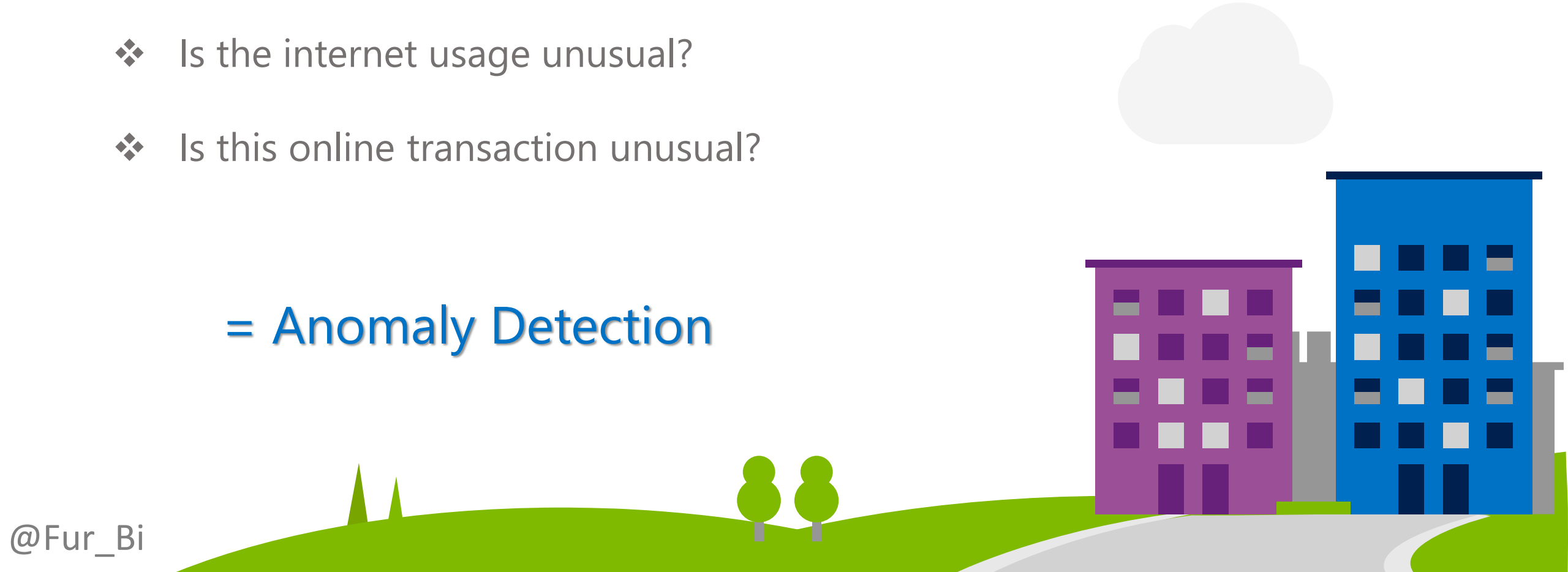
= Regression



Is this anomalous behaviour?

- Is this sensor reading out of the normal range?
- Is the internet usage unusual?
- Is this online transaction unusual?

= Anomaly Detection



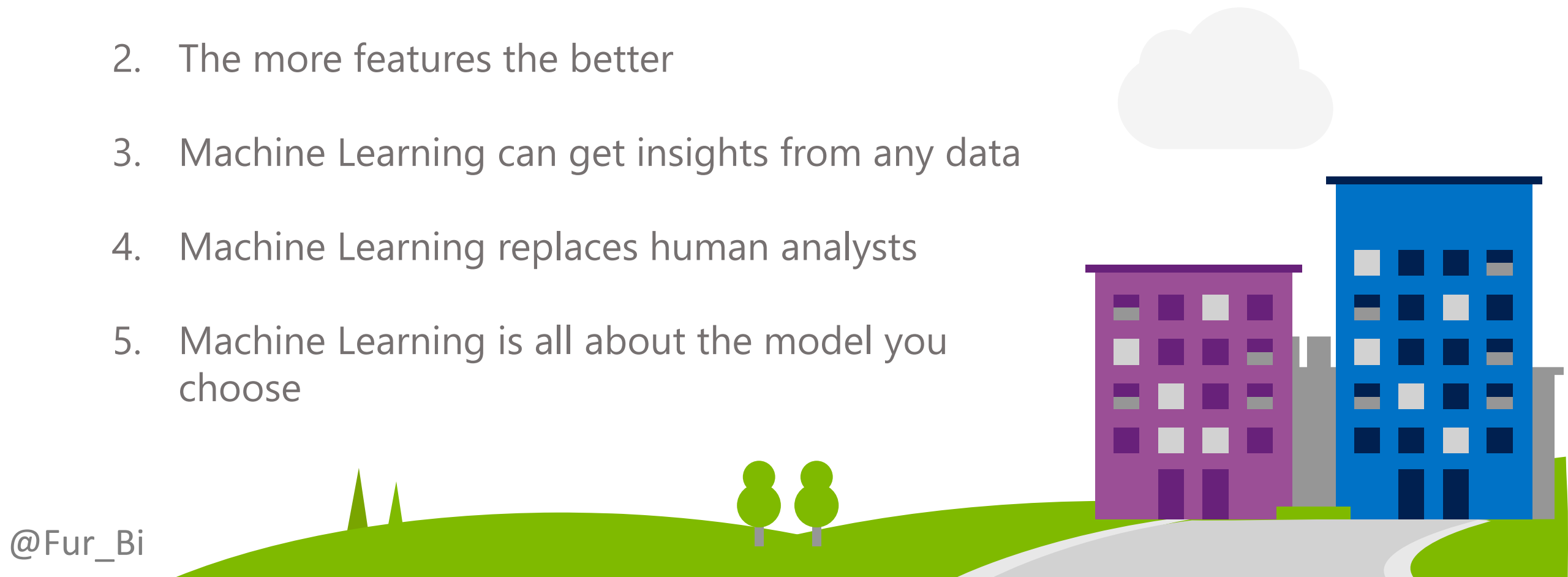
What are the patterns/groupings?

- Which of my customers have similar spending habits?
- How can these news articles be grouped?
- Which users have similar movie preferences?

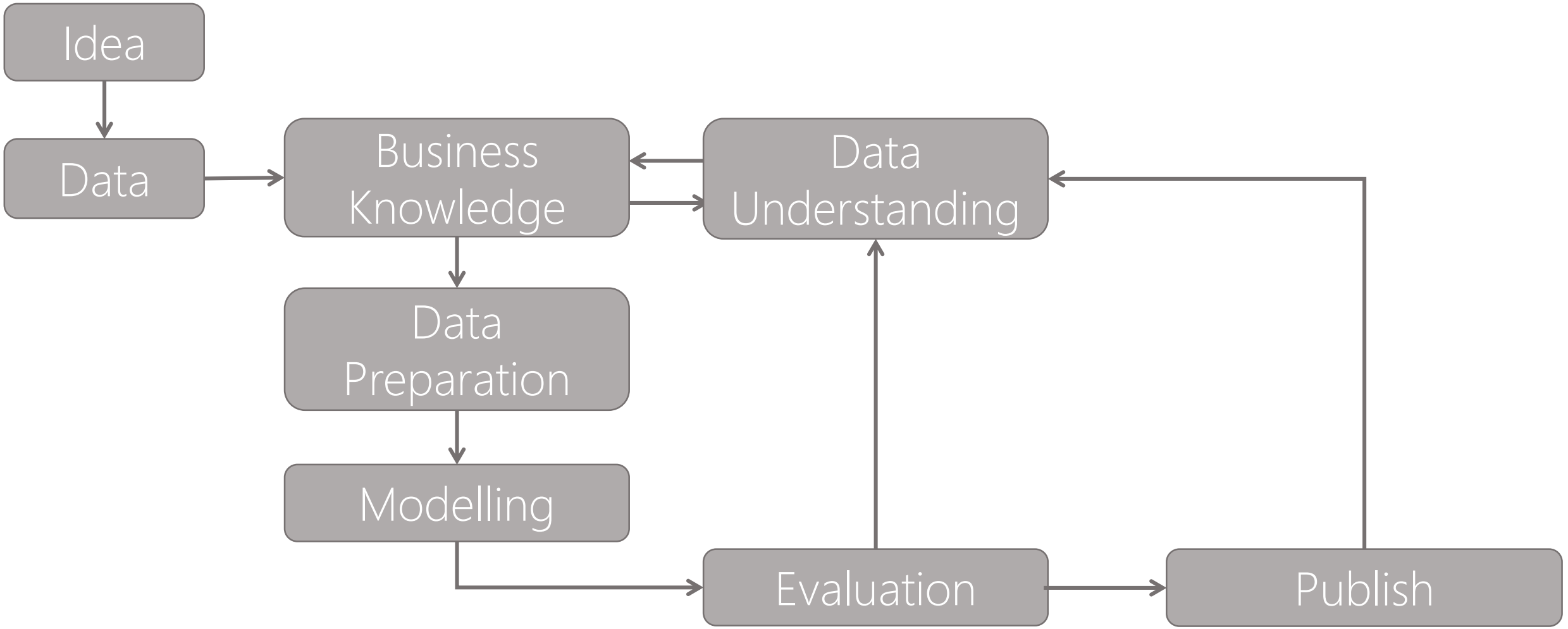
= Clustering

Machine Learning Myths

1. Machine Learning needs a lot of data
2. The more features the better
3. Machine Learning can get insights from any data
4. Machine Learning replaces human analysts
5. Machine Learning is all about the model you choose

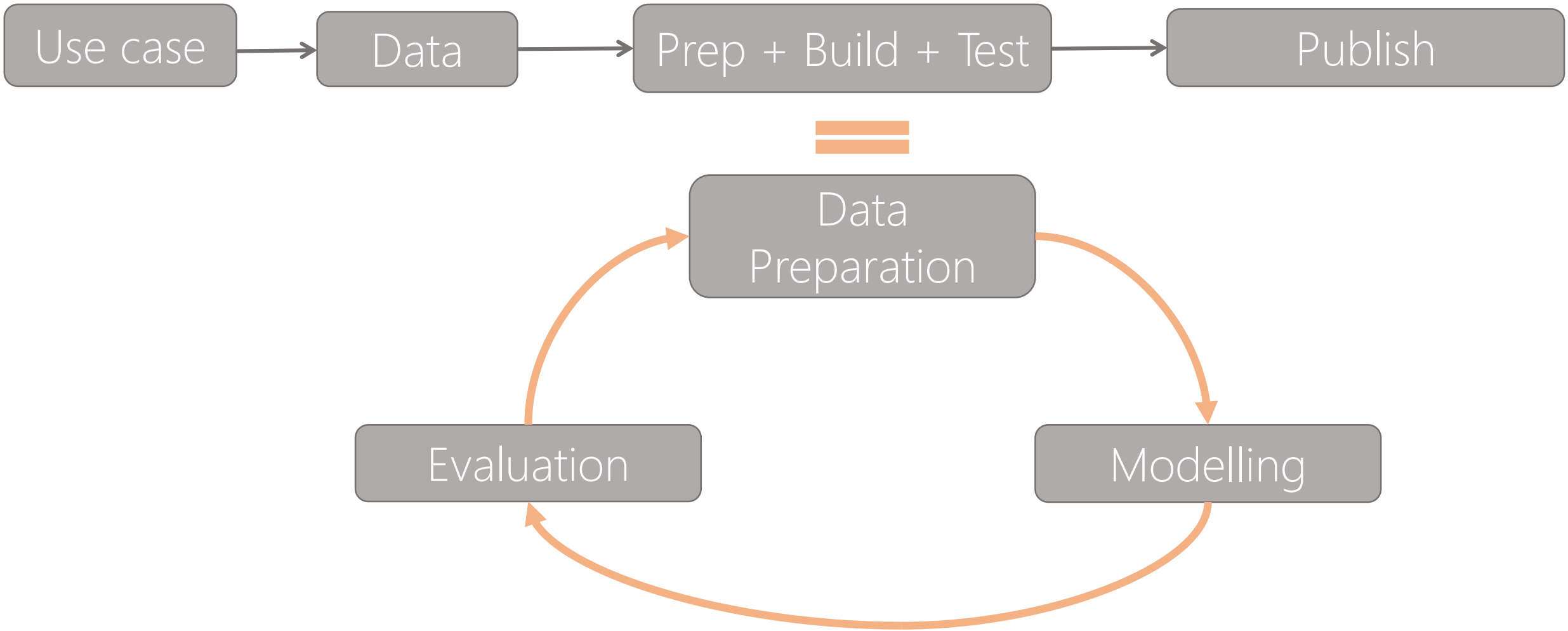


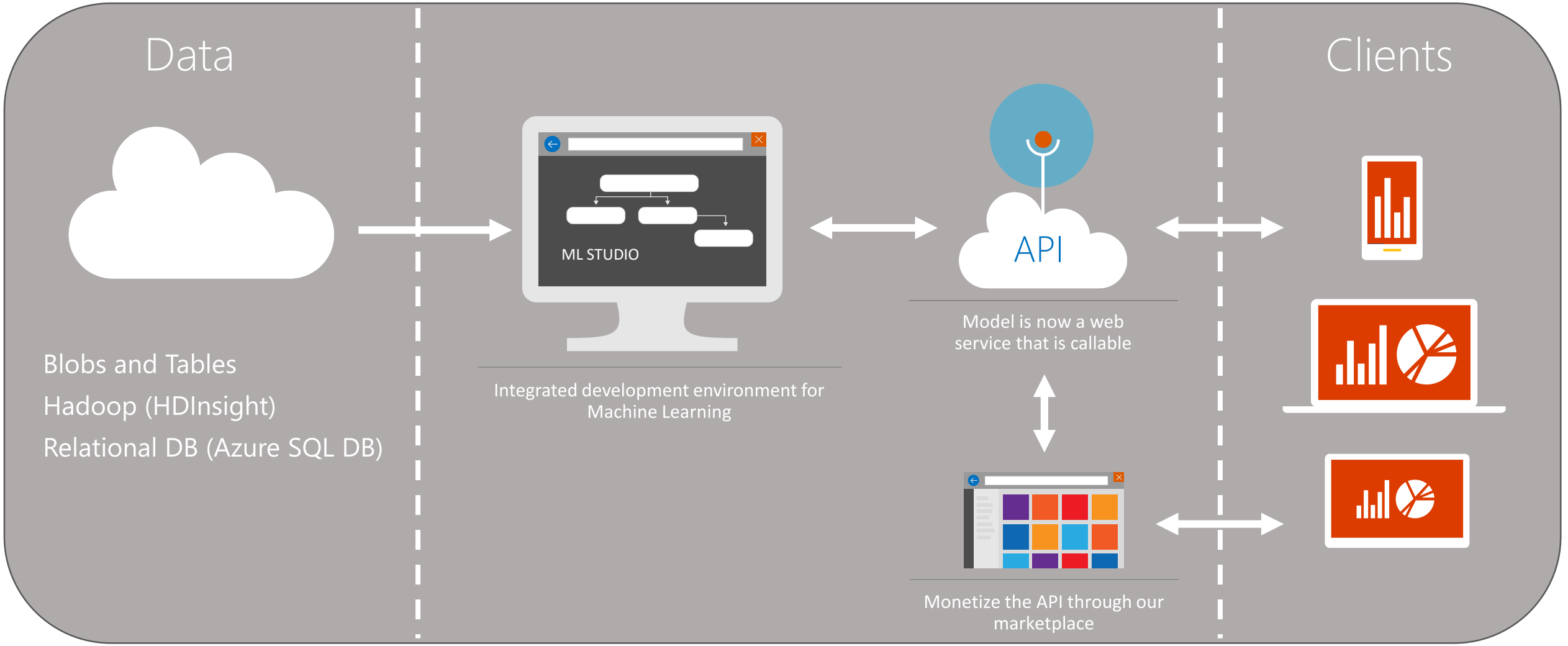
Machine Learning Process



Based on the [CRISP-DM Model]

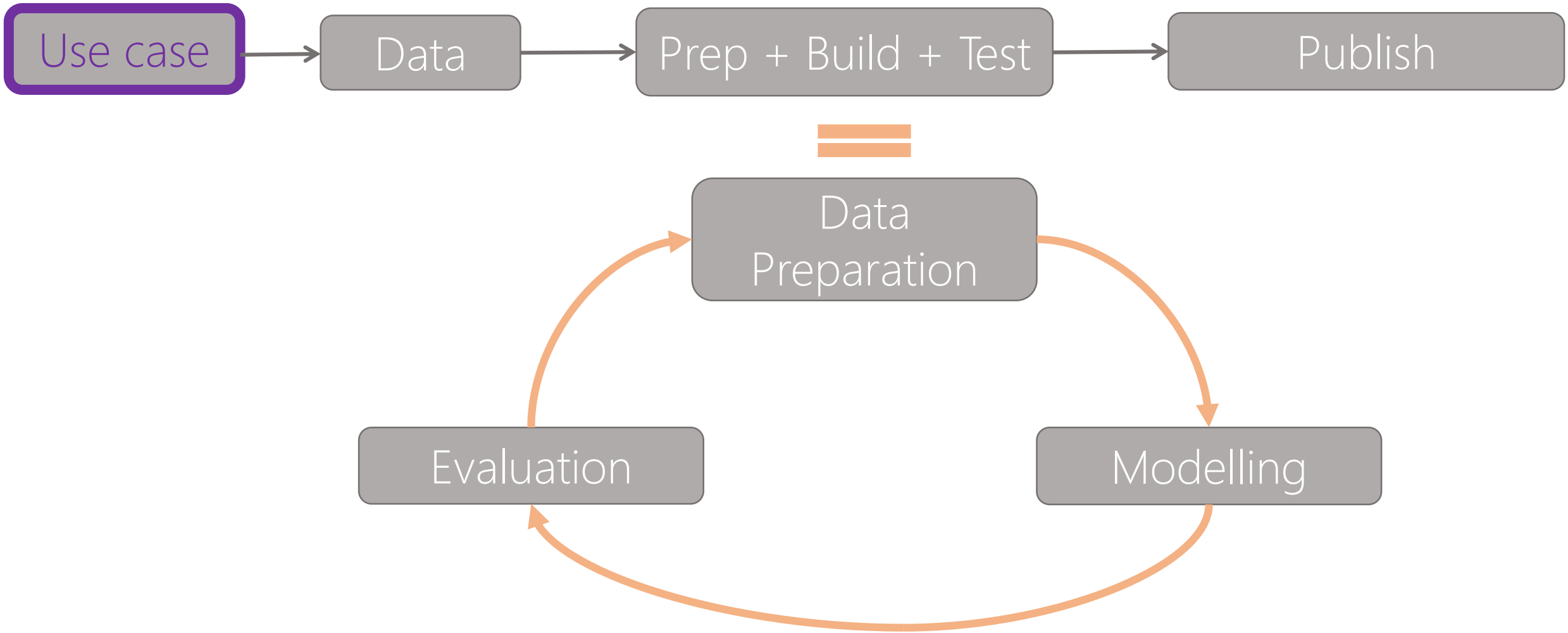
Machine Learning Process




Data
Clients
ML STUDIO
API
Blobs and Tables
Hadoop (HDInsight)
Relational DB (Azure SQL DB)
Integrated development environment for Machine Learning
Model is now a web service that is callable
Monetize the API through our marketplace

Demo

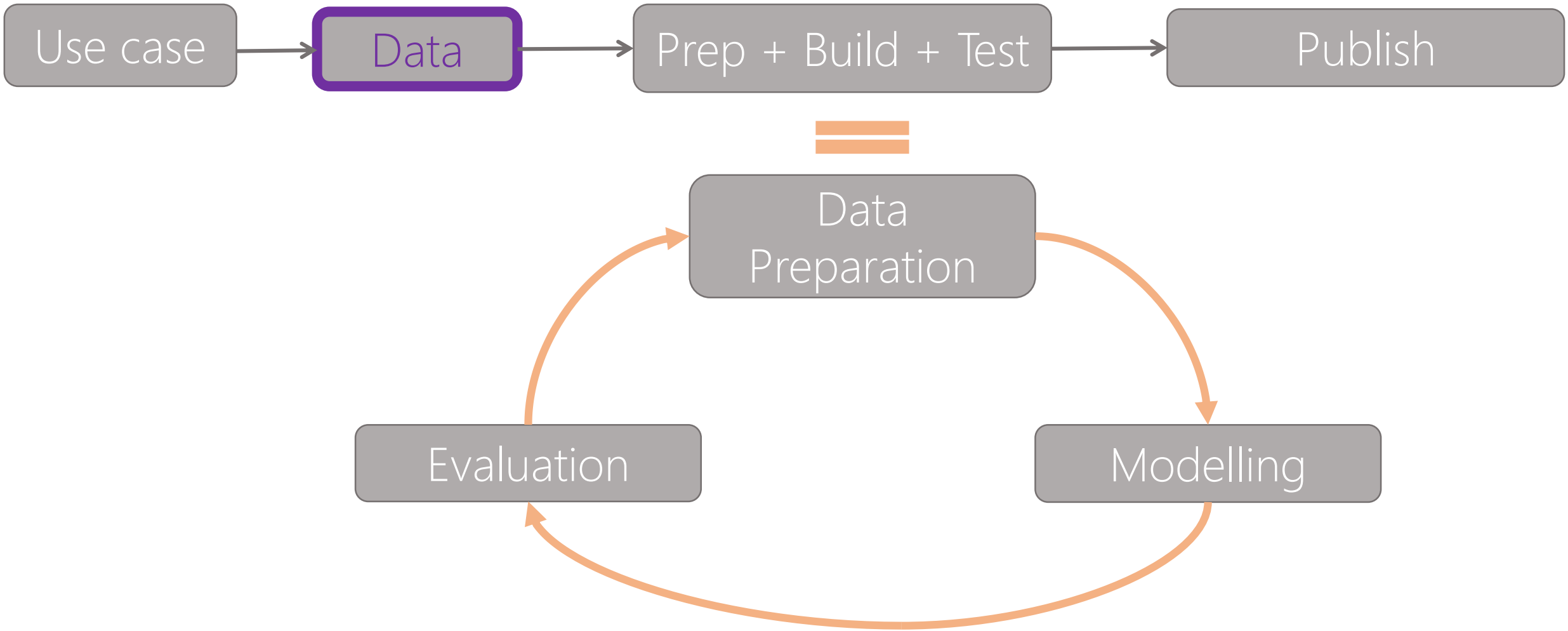
Machine Learning Process



Use Case/Scenario

AuthenticGiftsX is a company that is offering hand-made, personalised gifts which are dispatched in custom-made boxes. Their delivery labels are hand-written to give their customers a sense of authenticity. In order to improve and automate their delivery system they want a mobile app which can recognize the delivery address based on a picture of the label.

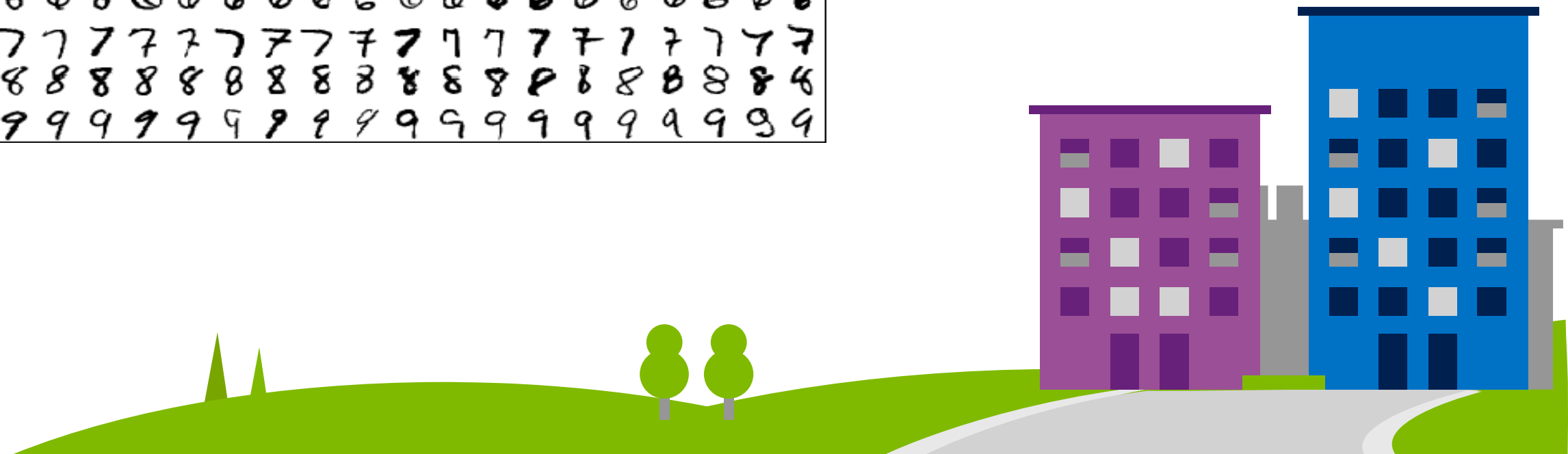
Machine Learning Process



Data

MNIST dataset
which consists of 70,000 grayscale images of hand-written digits

28x28 pixels



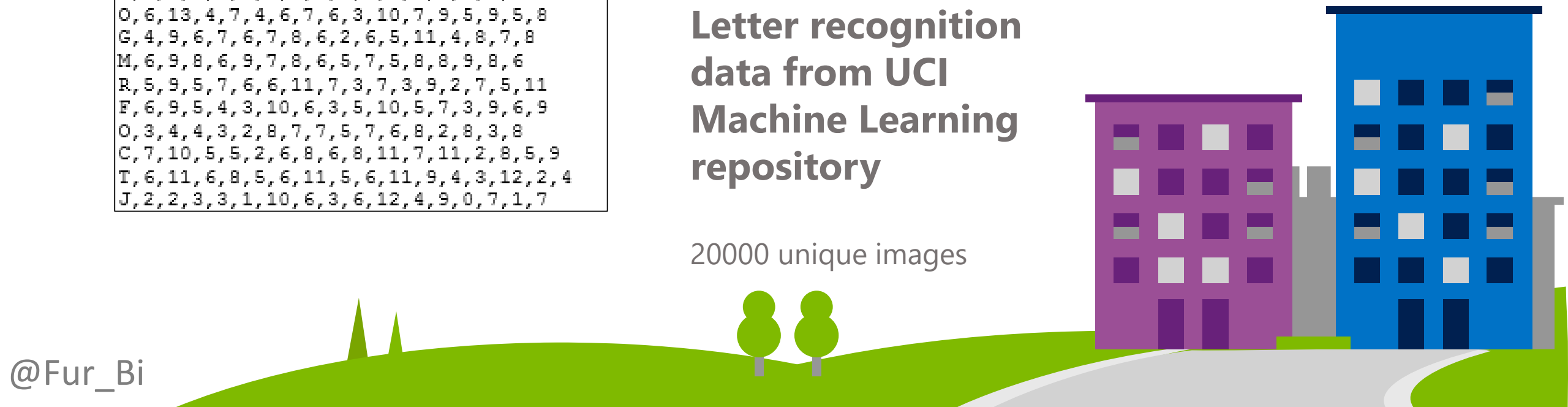
Data

```
T,2,8,3,5,1,8,13,0,6,6,10,8,0,8,0,8
I,5,12,3,7,2,10,5,5,4,13,3,9,2,8,4,10
D,4,11,6,8,6,10,6,2,6,10,3,7,3,7,3,9
N,7,11,6,6,3,5,9,4,6,4,4,10,6,10,2,8
G,2,1,3,1,1,8,6,6,6,6,5,9,1,7,5,10
S,4,11,5,8,3,8,8,6,9,5,6,6,0,8,9,7
B,4,2,5,4,4,8,7,6,6,7,6,6,2,8,7,10
A,1,1,3,2,1,8,2,2,2,8,2,8,1,6,2,7
J,2,2,4,4,2,10,6,2,6,12,4,8,1,6,1,7
M,11,15,13,9,7,13,2,6,2,12,1,9,8,1,1,8
X,3,9,5,7,4,8,7,3,8,5,6,8,2,8,6,7
O,6,13,4,7,4,6,7,6,3,10,7,9,5,9,5,8
G,4,9,6,7,6,7,8,6,2,6,5,11,4,8,7,8
M,6,9,8,6,9,7,8,6,5,7,5,8,8,9,8,6
R,5,9,5,7,6,6,11,7,3,7,3,9,2,7,5,11
F,6,9,5,4,3,10,6,3,5,10,5,7,3,9,6,9
O,3,4,4,3,2,8,7,7,5,7,6,8,2,8,3,8
C,7,10,5,5,2,6,8,6,8,11,7,11,2,8,5,9
T,6,11,6,8,5,6,11,5,6,11,9,4,3,12,2,4
J,2,2,3,3,1,10,6,3,6,12,4,9,0,7,1,7
```

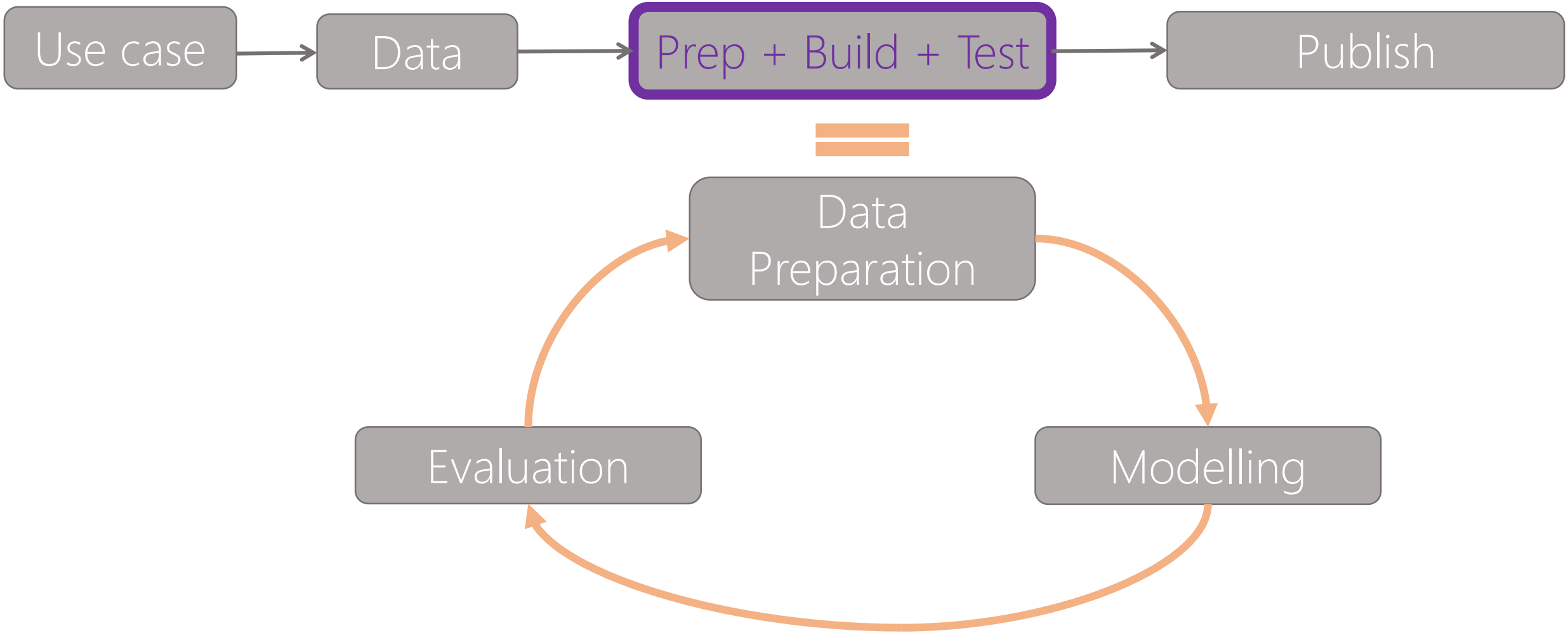
```
Attribute Information:
      1.     lettr    capital letter  (26 values from A to Z)
      2.     x-box    horizontal position of box      (integer)
      3.     y-box    vertical position of box        (integer)
      4.     width    width of box                    (integer)
      5.     high     height of box                   (integer)
      6.     onpix    total # on pixels               (integer)
      7.     x-bar    mean x of on pixels in box      (integer)
      8.     y-bar    mean y of on pixels in box      (integer)
      9.     x2bar    mean x variance                 (integer)
     10.     y2bar    mean y variance                 (integer)
     11.     xybar    mean x y correlation            (integer)
     12.     x2ybr    mean of x * x * y               (integer)
     13.     xy2br    mean of x * y * y               (integer)
     14.     x-ege    mean edge count left to right   (integer)
     15.     xegvy    correlation of x-ege with y     (integer)
     16.     y-ege    mean edge count bottom to top   (integer)
     17.     yegvx    correlation of y-ege with x     (integer)
```

Letter recognition data from UCI Machine Learning repository

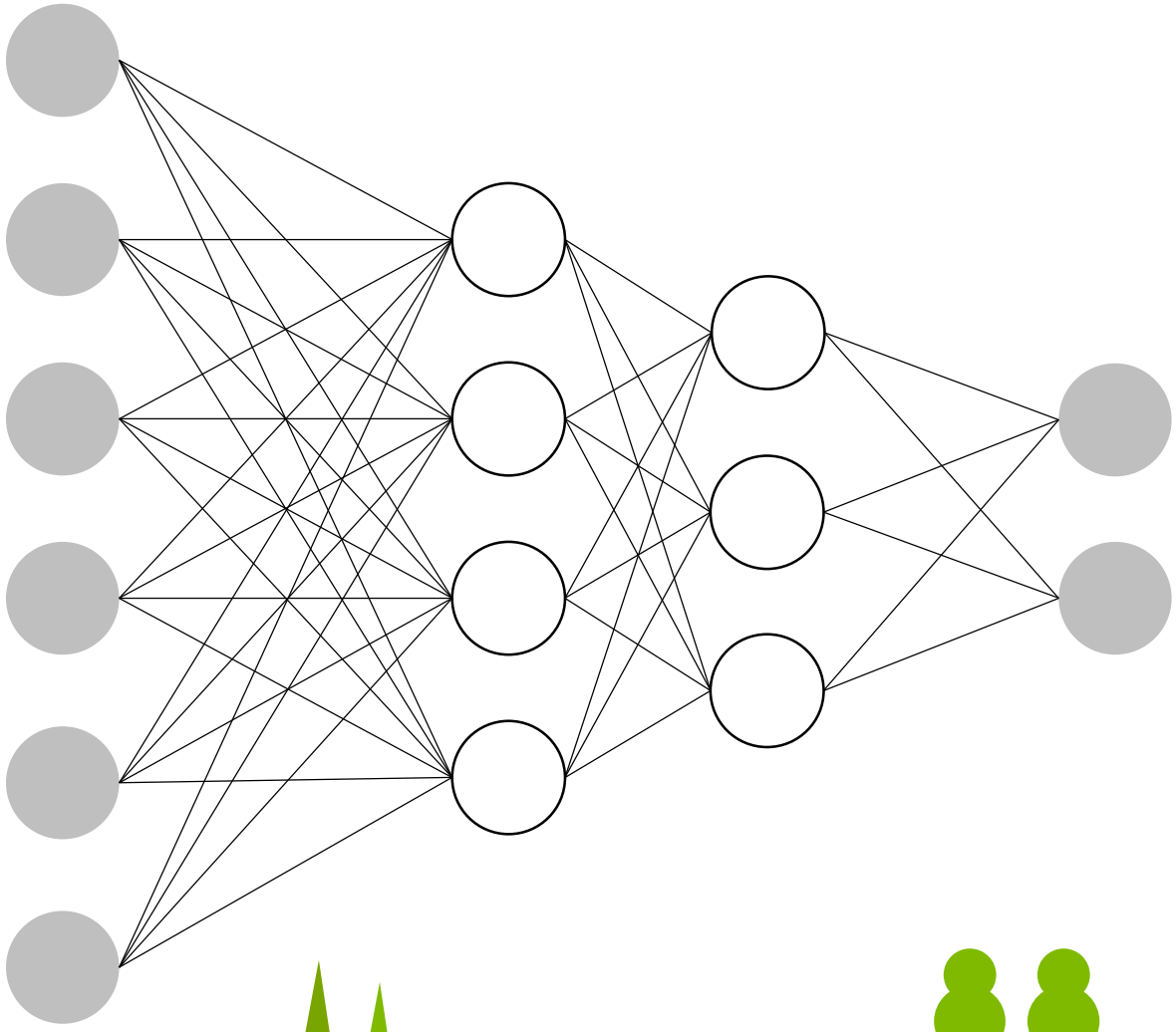
20000 unique images



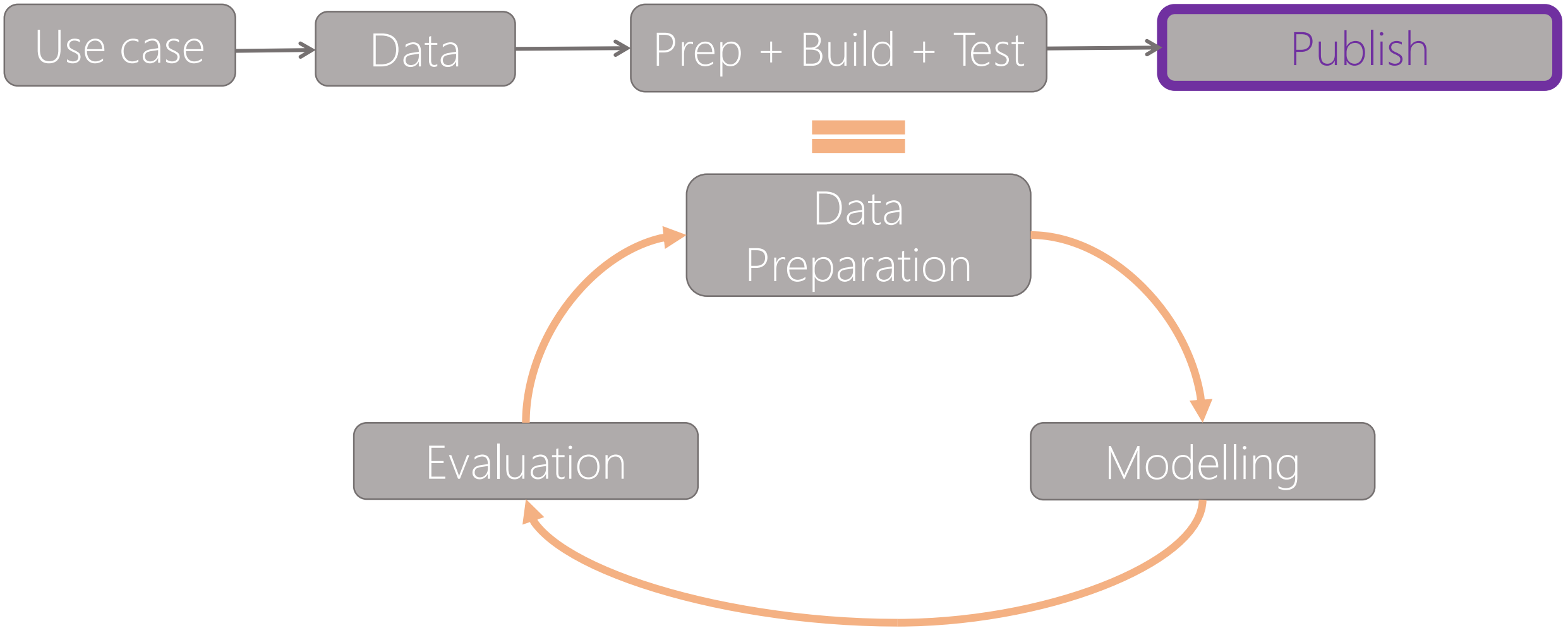
Machine Learning Process



What is a neural network?

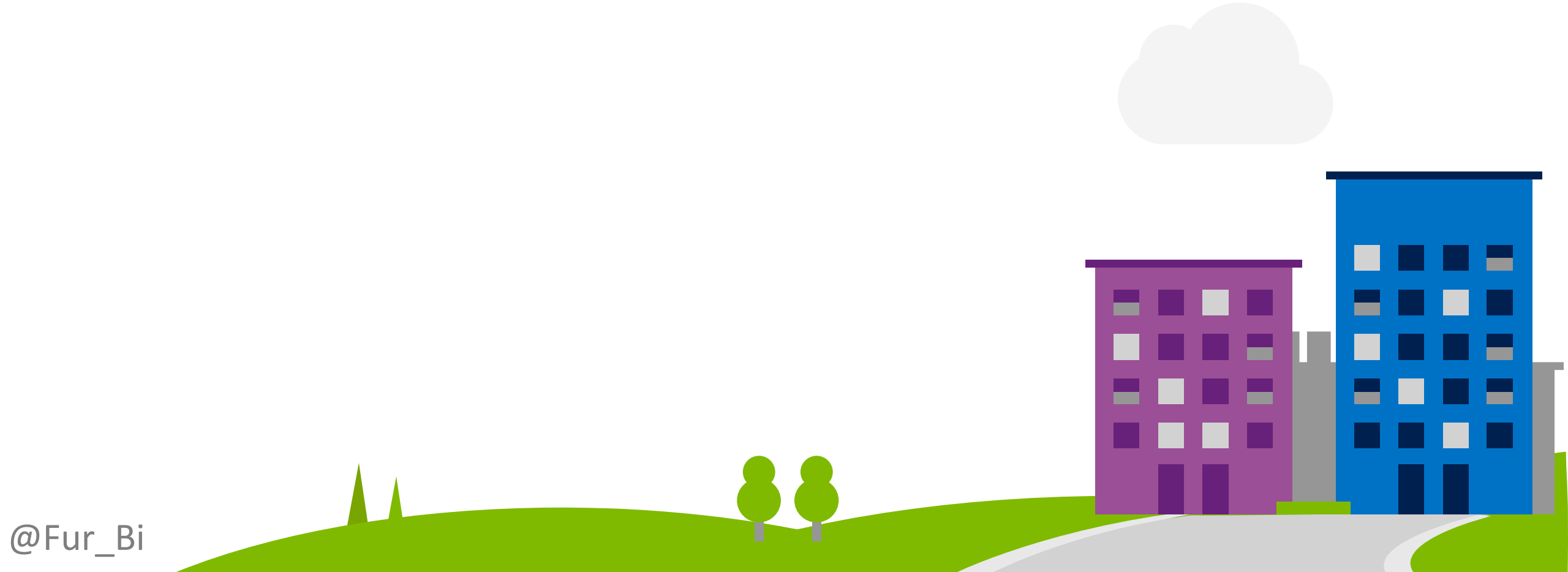


Machine Learning Process

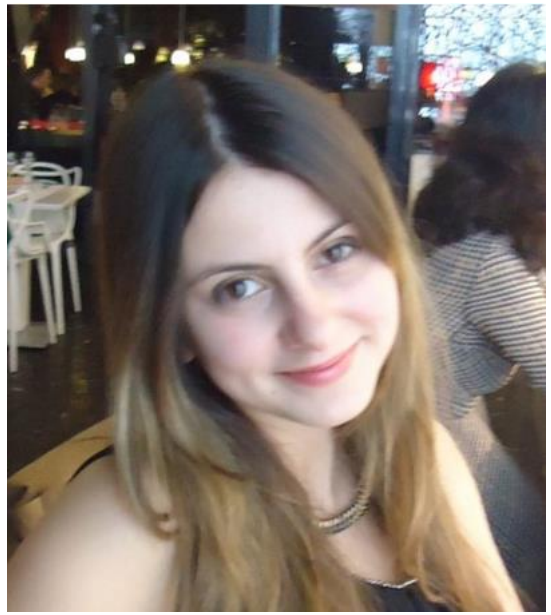


Action !!!

Go to Azure Machine Learning Studio and create your first experiment.



Thank you!



@Fur_Bi

bifurt@microsoft.com

http://furbi.azurewebsites.net/