Machine Learning and AIOps

Splunk4Ninjas | [CUSTOMER NAME]

.conf24

splunk>

Forward-looking statements

This presentation may contain forward-looking statements that are subject to the safe harbors created under the Securities Act of 1933, as amended, and the Securities Exchange Act of 1934, as amended. All statements other than statements of historical facts are statements that could be deemed forward-looking statements. These statements are based on current expectations, estimates, forecasts, and projections about the industries in which we operate and the beliefs and assumptions of our management based on the information currently available to us. Words such as “expects,” “anticipates,” “targets,” “goals,” “projects,” “intends,” “plans,” “believes,” “momentum,” “seeks,” “estimates,” “continues,” “endeavors,” “strives,” “may,” variations of such words, and similar expressions are intended to identify such forward-looking statements. In addition, any statements that refer to (1) our goals, commitments, and programs; (2) our business plans, initiatives, and objectives; and (3) our assumptions and expectations, including our expectations regarding our financial performance, products, technology, strategy, customers, markets, acquisitions and investments are forward-looking statements. These forward-looking statements are not guarantees of future performance and involve significant risks, uncertainties and other factors that may cause our actual results, performance or achievements to be materially different from results, performance or achievements expressed or implied by the forward-looking statements contained in this presentation. Readers are cautioned that these forward-looking statements are only predictions and are subject to risks, uncertainties, and assumptions that are difficult to predict, including those identified in the “Risk Factors” section of Cisco’s most recent report on Form 10-Q filed on May 21, 2024 and its most recent report on Form 10-K filed on September 7, 2023. The forward-looking statements made in this presentation are made as of the time and date of this presentation. If reviewed after the initial presentation, even if made available by Cisco or Splunk, on Cisco or Splunk’s website or otherwise, it may not contain current or accurate information. Cisco and Splunk undertake no obligation to revise or update any forward-looking statements for any reason, except as required by law.

In addition, any information about new products, features, functionality or our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment or be relied upon in making a purchasing decision. We undertake no commitment, promise or obligation either to develop the features or functionalities described, in beta or in preview (used interchangeably), or to include any such feature or functionality in a future release. The development, release, and timing of any features or functionality described for our products remains at our sole discretion.

Meet Your Presenters

[NAME]
[Title]

[NAME]
[Title]

[NAME]
[Title]

Agenda

Current Challenges for Operations Teams

How Splunk Drives Machine Learning

Scenario Introduction

Deep Dive Modules

Wrap Up and Next Steps

Leading Initiatives Driving ML Adoption

200%

Increase in proactive detection of security and performance issues, significantly reducing downtime

2.1x

More likely to have automated processes for alerts, helping operationalize data at scale

$365k/hour

On average saved from costly outages, helping organizations protect against revenue loss

Sources:
Harvard Business Review - Artificial Intelligence for the Real World Digital
Enterprise Journal Report: The Roadmap to Becoming a Top Performing Organization in Managing IT Operations

Obstacles Blocking ML Adoption

1.8x

Increase in data and events to process every two years, creating challenges in handling data volume

1 in 2

Companies increase the number of data silos, leading to difficulties integrating ML in isolated systems

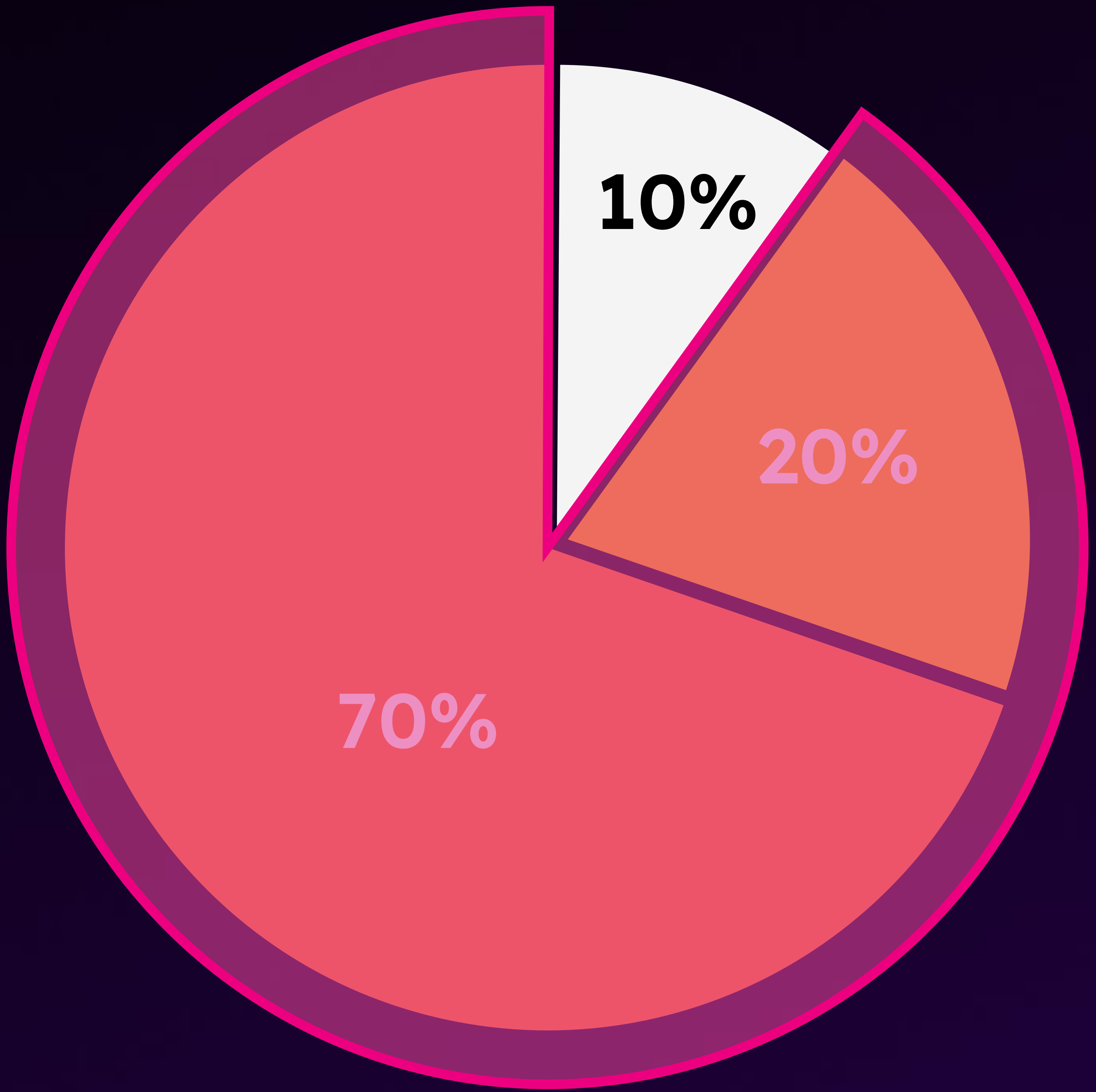
79%

Failure rate for companies which try to implement machine learning from scratch, due to lack of expertise

Sources:
Harvard Business Review - Artificial Intelligence for the Real World
Digital Enterprise Journal Report: The Roadmap to Becoming a Top Performing Organization in Managing IT Operations

https://www.xplm.com/news/press/industry-study-2023-companies-cannot-control-their-data-silos/

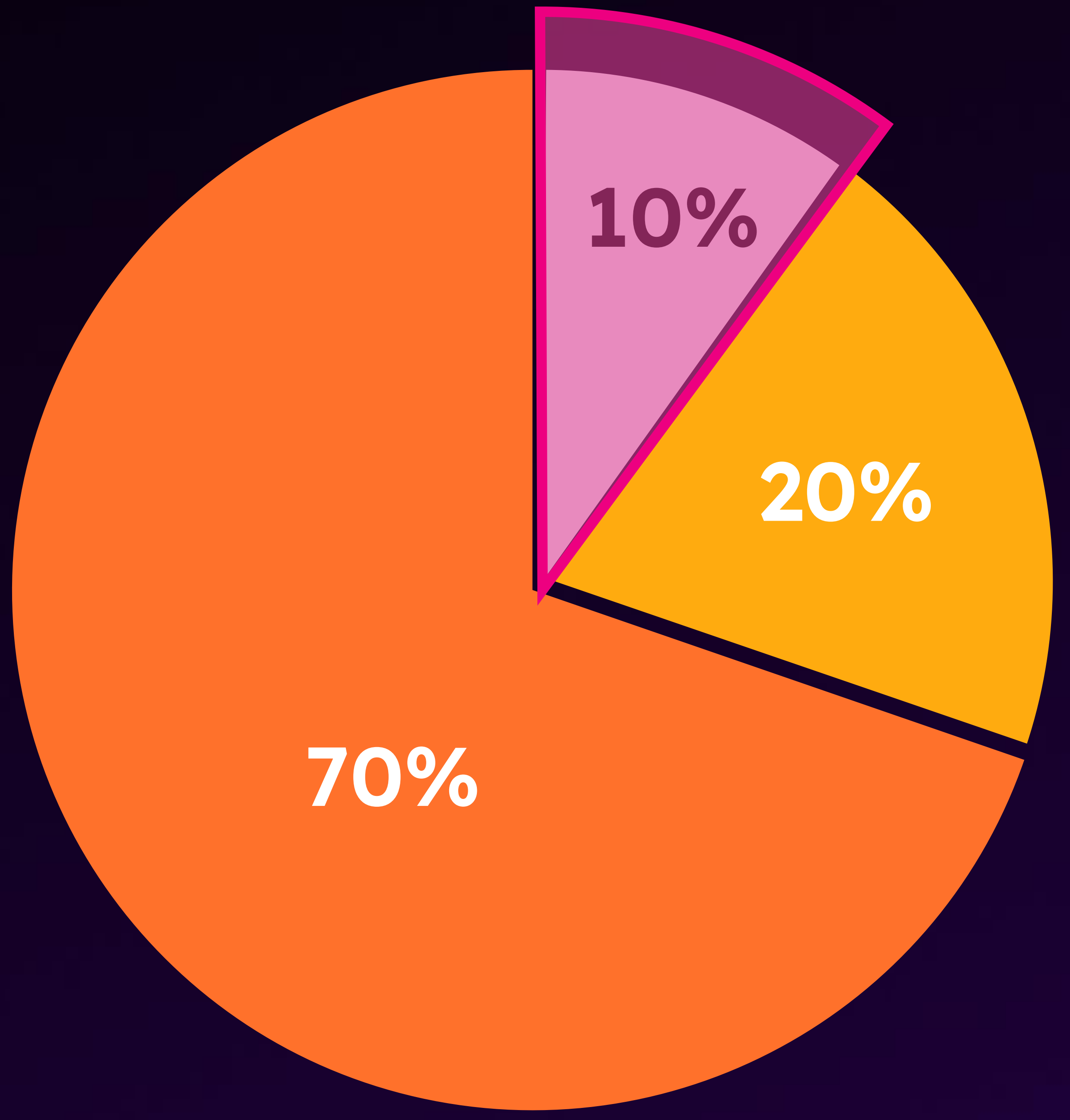
How Data Scientists Spend Their Time



- Data Engineering
- Machine Learning
- Other

Data originated from "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", Forbes Mar 23, 2016".

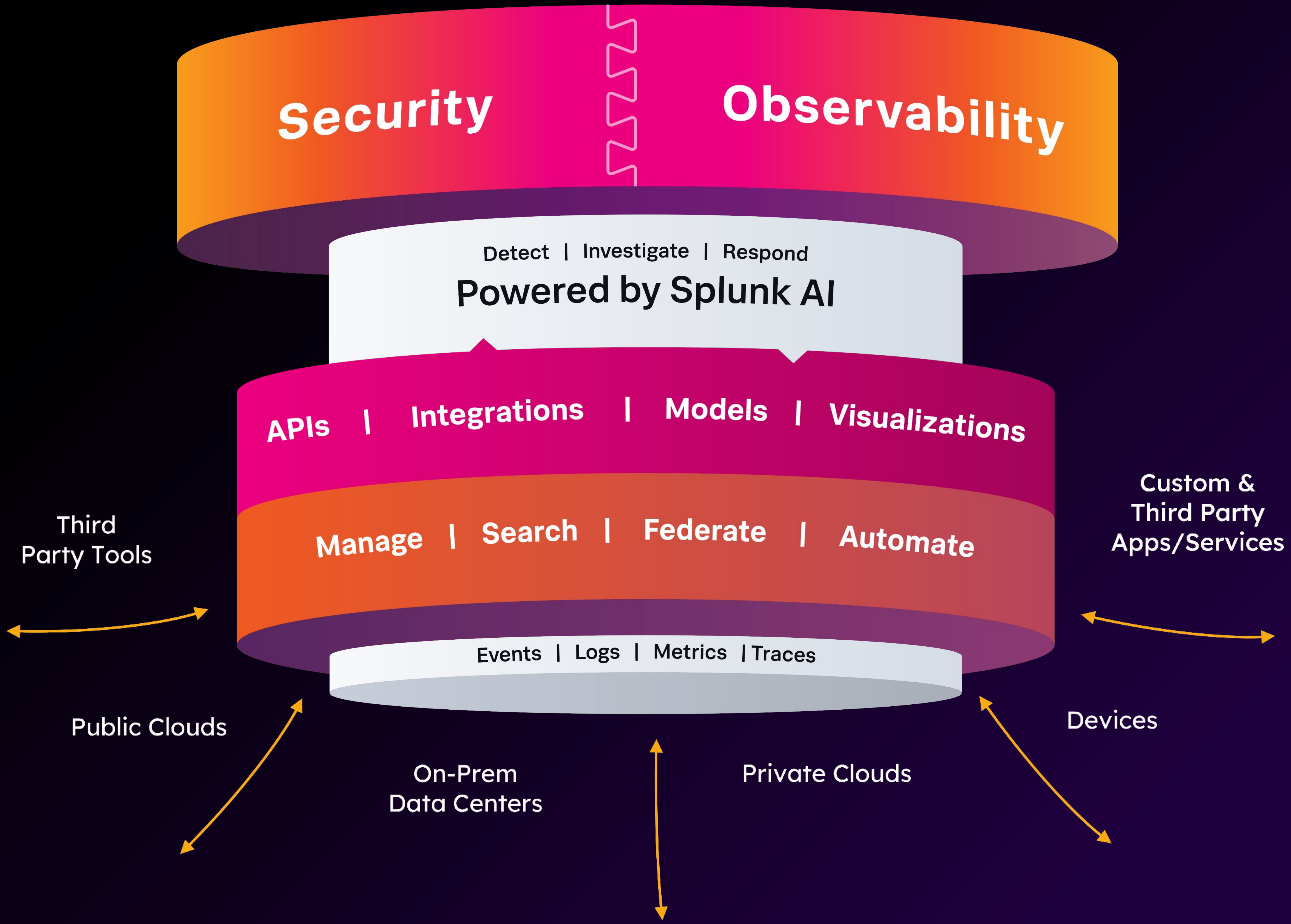
How Data Scientists Spend Their Time



- Data Engineering
- Machine Learning
- Other

Data originated from "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", Forbes Mar 23, 2016".

The Unified Security and Observability Platform



Different Levels of Teaching Machines

Artificial Intelligence

The broad study of teaching a computer to process data and make decisions

Different Levels of Teaching Machines

Artificial Intelligence

The broad study of teaching a computer to process data and make decisions

Machine Learning

Subset of AI. Predictions and insight with minimal human interference

Different Levels of Teaching Machines

Artificial Intelligence

The broad study of teaching a computer to process data and make decisions

Machine Learning

Subset of AI. Predictions and insight with minimal human interference

Deep Learning

Subset of ML. Predictions via neural networks

Different Levels of Teaching Machines

Artificial Intelligence

The broad study of teaching a computer to process data and make decisions

Machine Learning

Subset of AI. Predictions and insight with minimal human interference

Deep Learning

Subset of ML. Predictions via neural networks

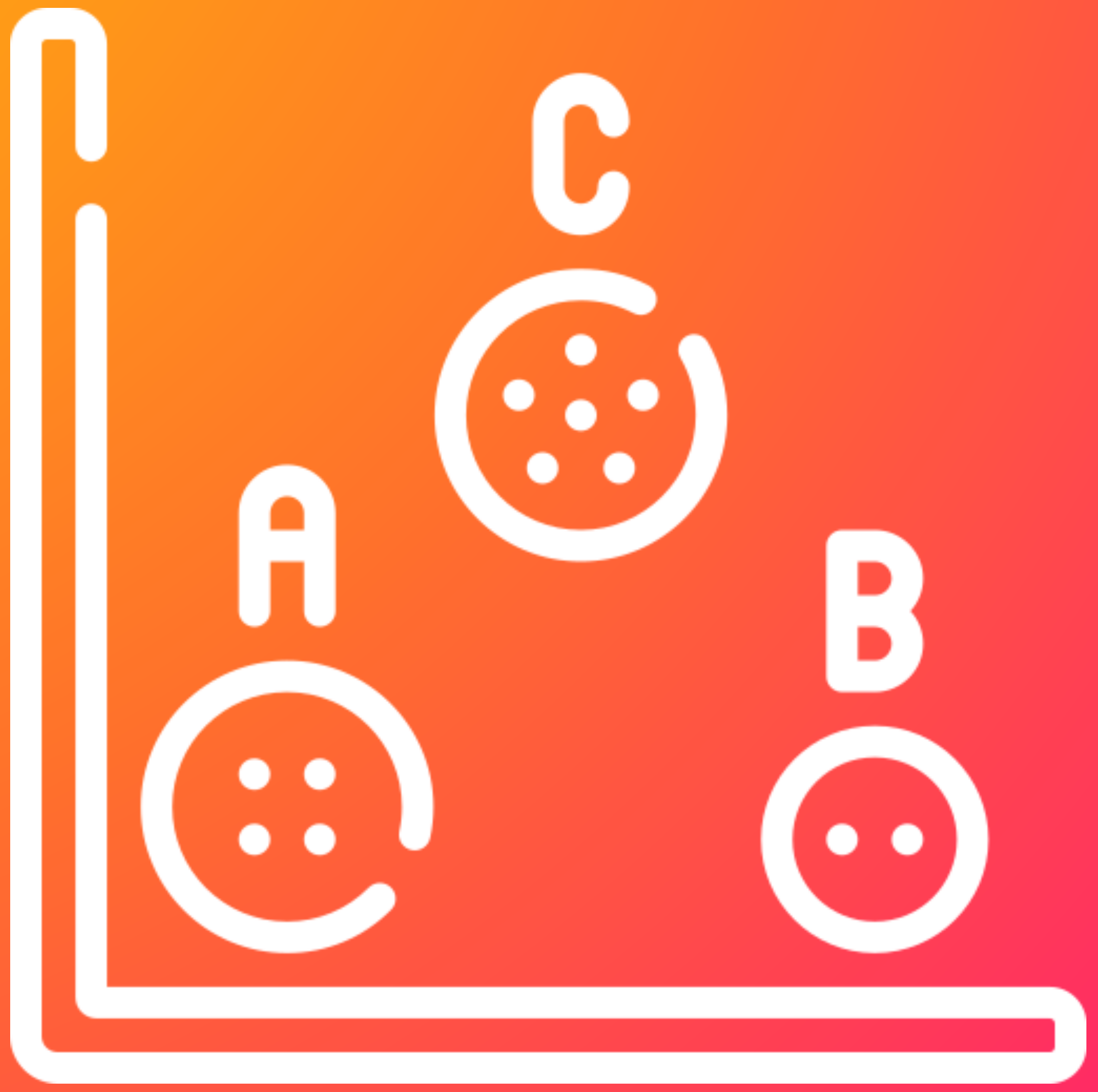
Predictive Algorithms

Methods that help you get ahead of issues that may happen in the future

Includes:

- Numerical Regression
- Categorical Regression
- Time Series Forecasting

Source: Parzival 1997, flaticon.com



Categorization Algorithms

Uncover insights about your data to quickly respond in the present

Includes:

- Categorical Regression
- Clustering

Source:Freepik, flaticon.com

Outlier Detection Algorithms

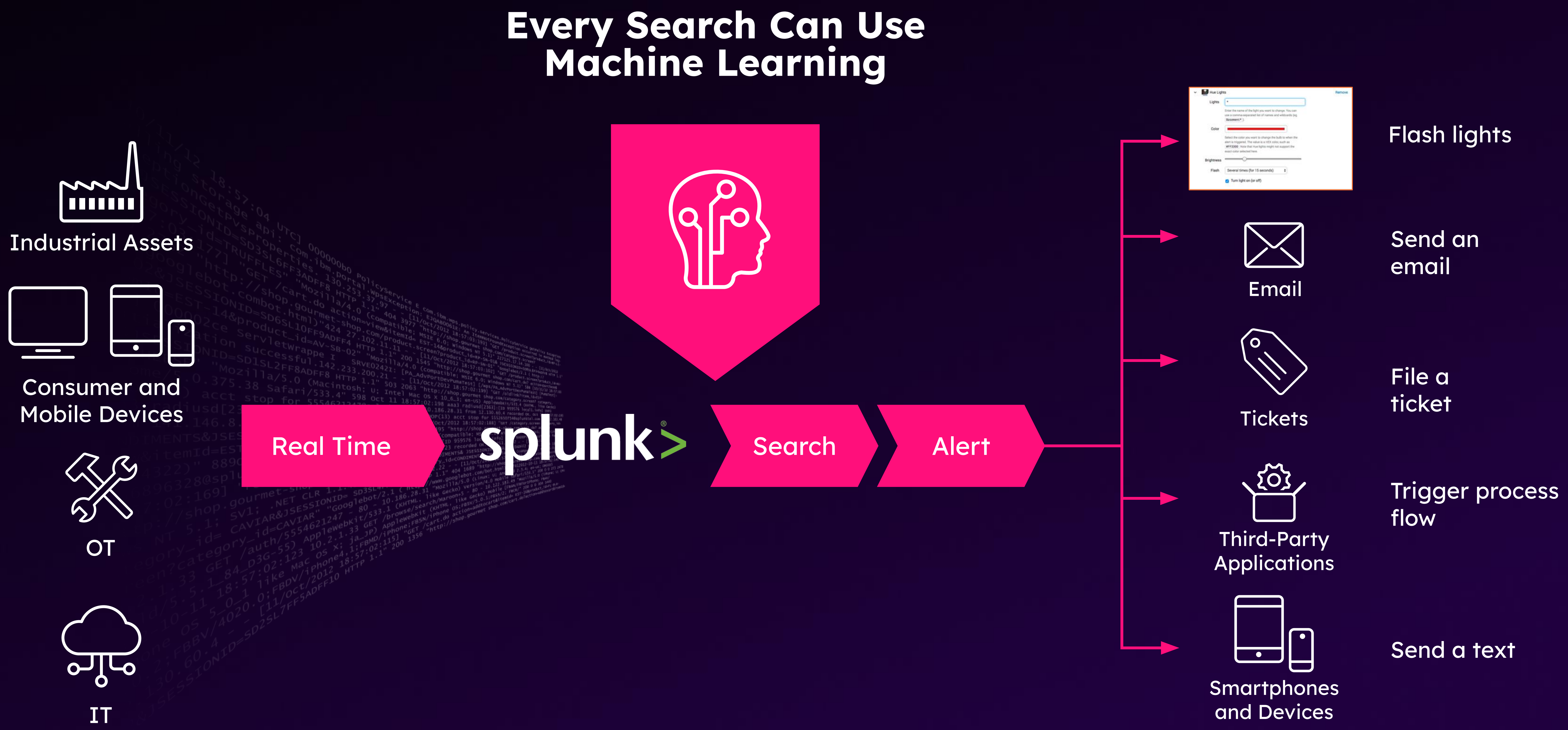
Identify and analyze abnormal behavior in your data

Includes:

- Clustering
- Outlier Detection

Source:Hajicon, flaticon.com

Easy to Operationalize



Model Longevity

NUMERICAL REGRESSION START

Use Case:

Modeling System Behavior

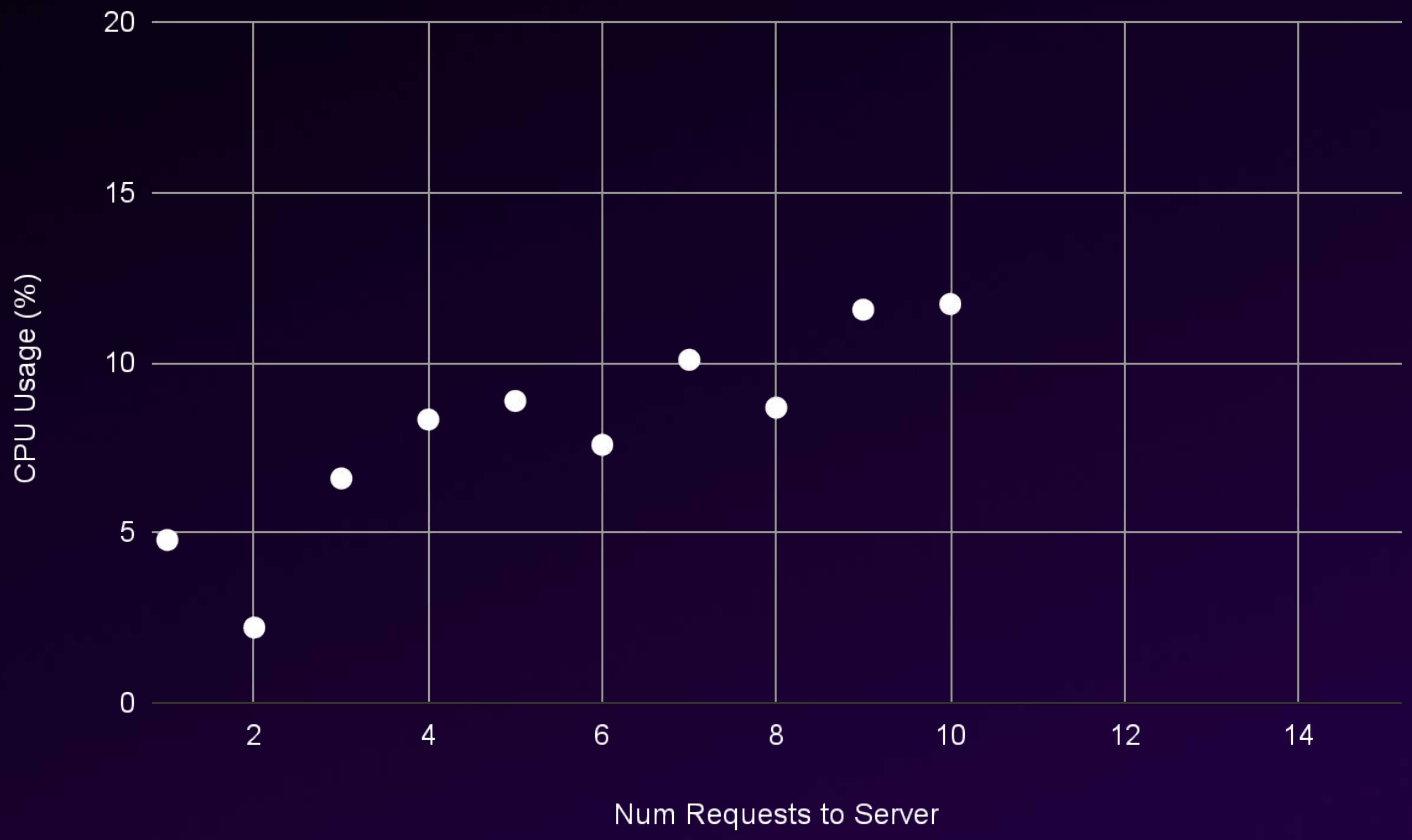
Source:Freepik, flaticon.com

Numeric Prediction

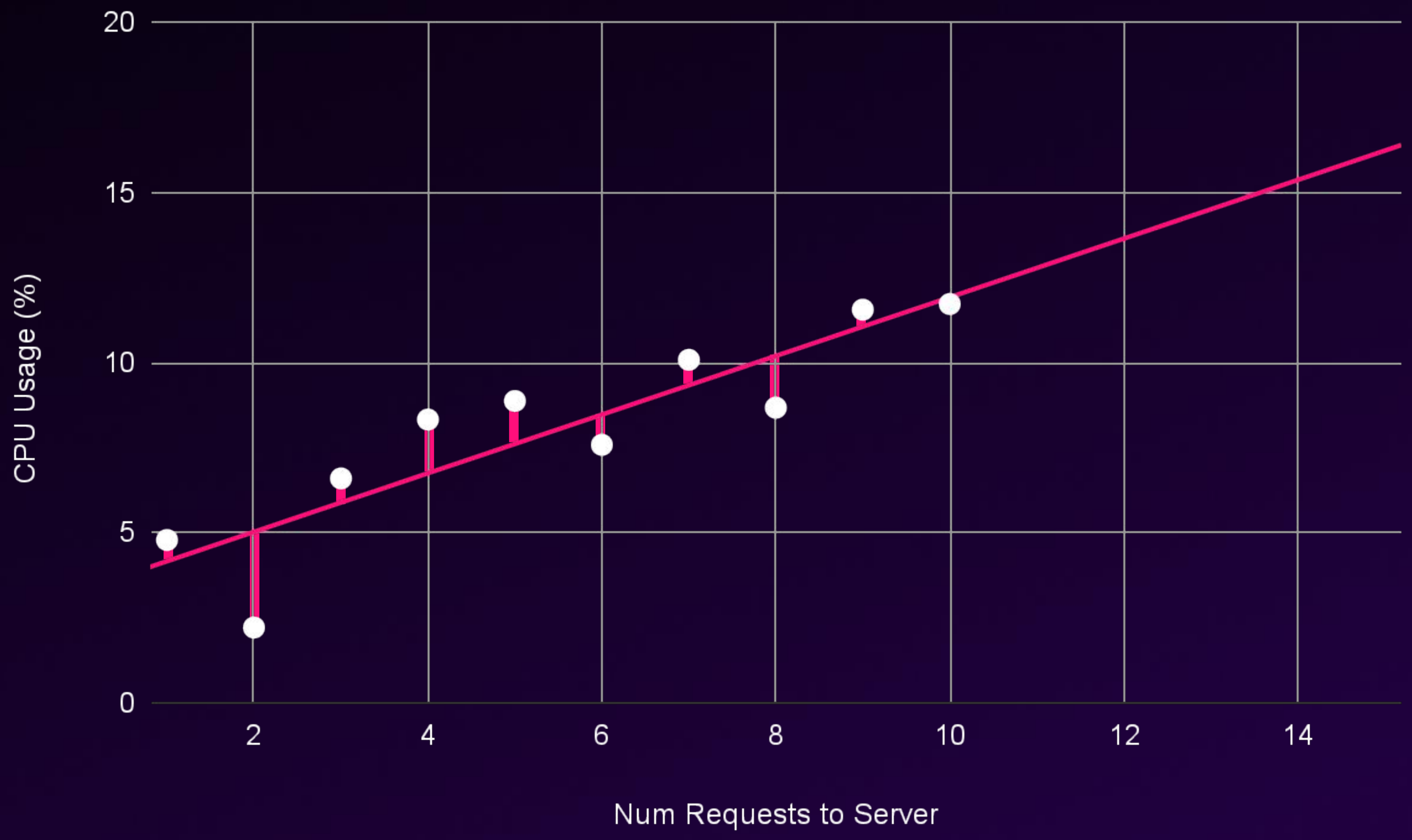
"A method that lets you model and predict how a metric will behave based on changes in the environment"

Source:Becris, flaticon.com

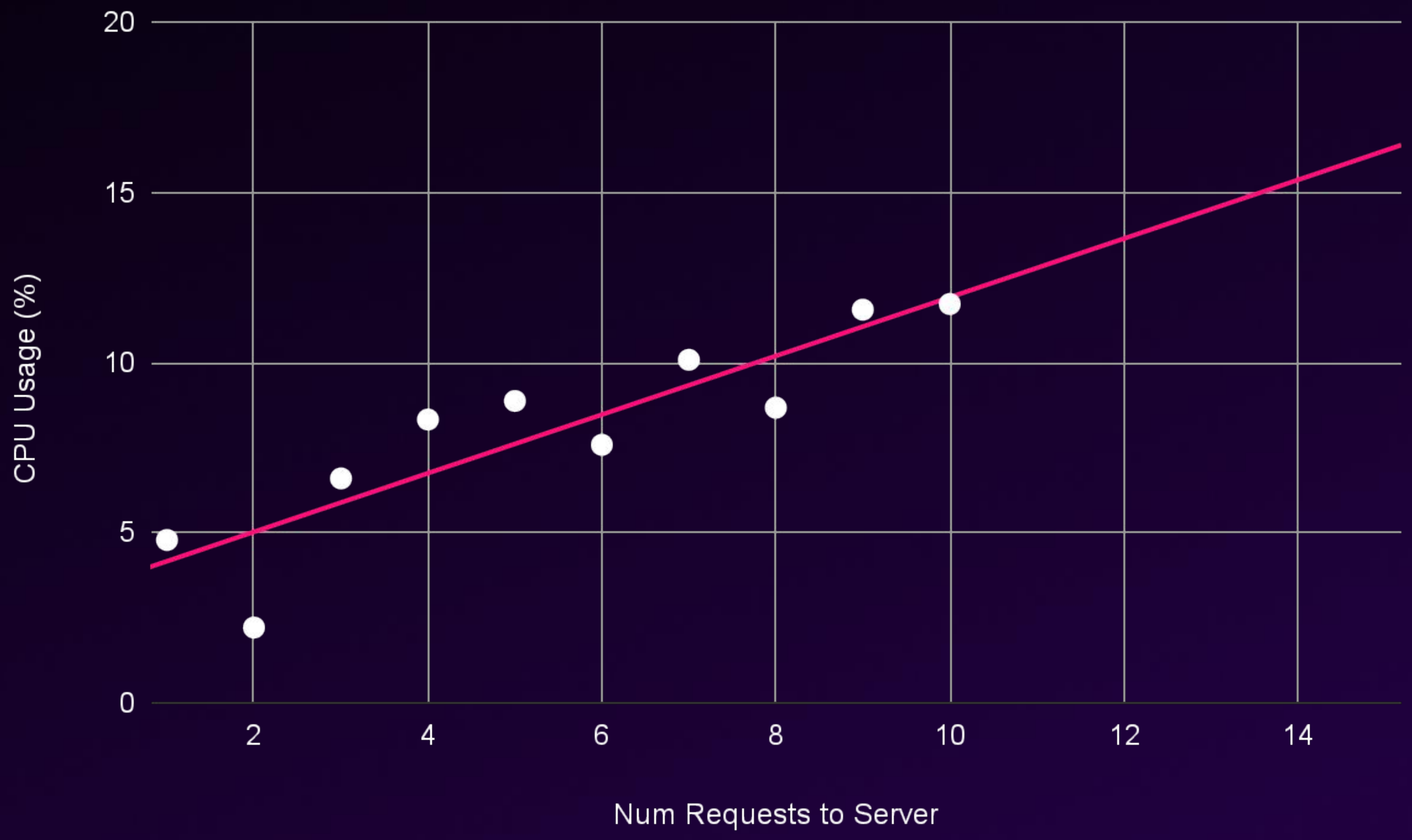
Numerical Prediction



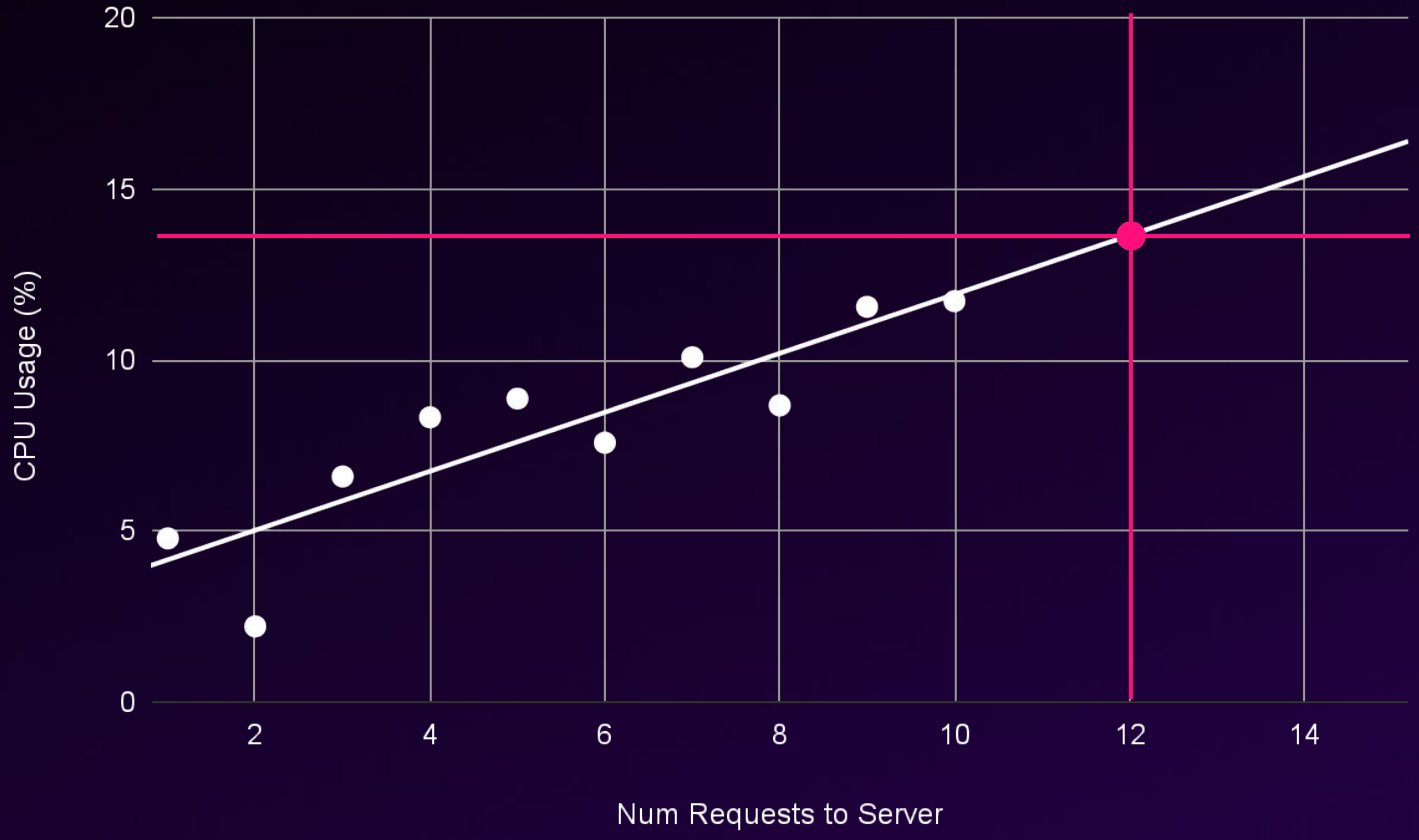
Numerical Prediction



Numerical Prediction



Numerical Prediction



Numeric Prediction

"A method that lets you model and predict how a metric will behave based on changes in the environment"

Live Instance Demo

Log Into [INSTANCE URL]

Lab Guide Exercise #1

Time: 10 minutes

Summary

Top 4 most important things to remember about numerical prediction

1

Predicting numeric fields is done using a **supervised learning** method which uses **labeled data**

2

Models **assumes a causative relationship exists** among selected fields

3

Scaling data prior to training is **almost always necessary**

4

Choice of numeric prediction algorithm(s) may rely on a **subject matter expert** of the data

Predictive Analytics

NUMERICAL REGRESSION END

FORECASTING START

Use Case:

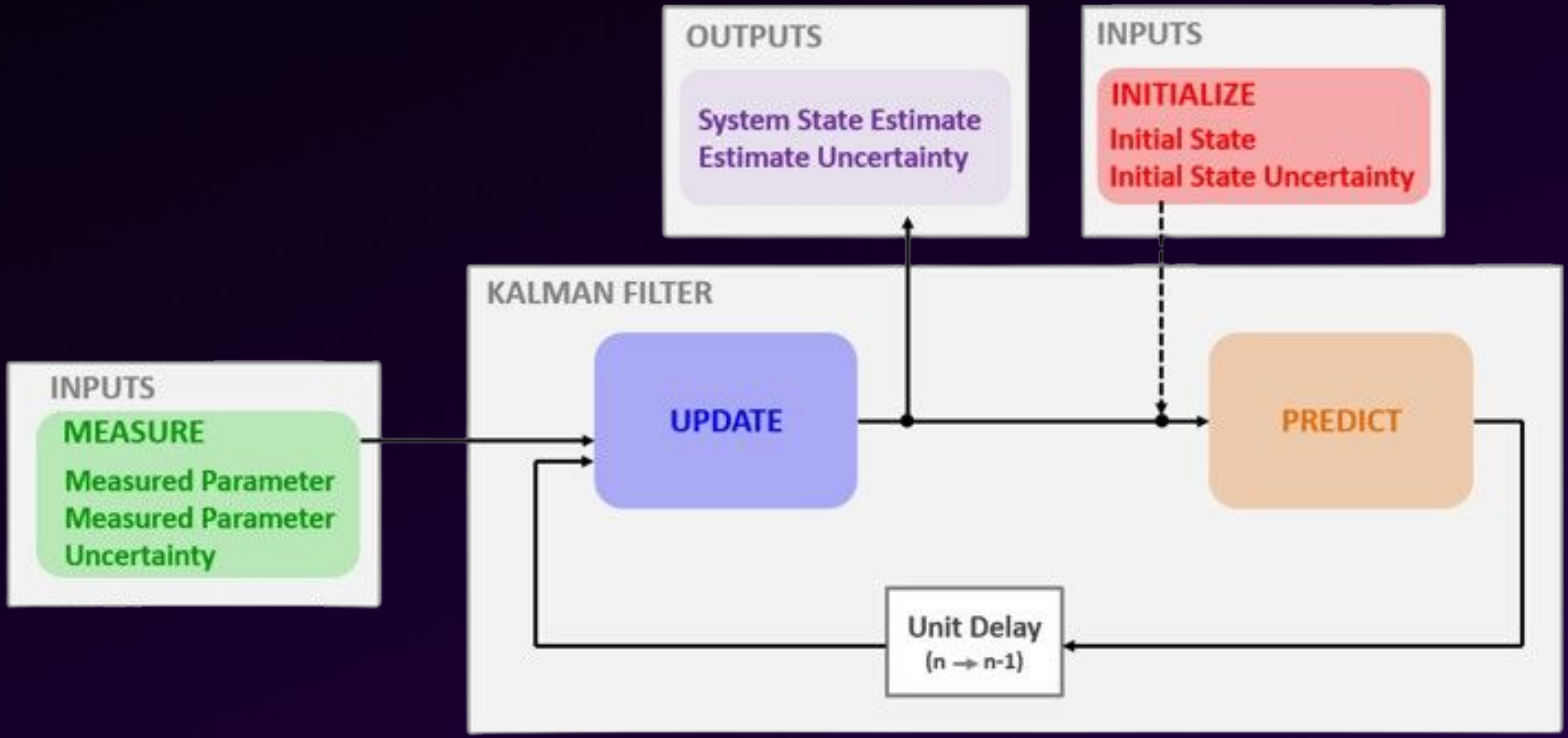
Forecasting Key Metrics

Forecasting Time Series

“Using historical data to identify patterns, which are then used to forecast how your data might behave in the future”

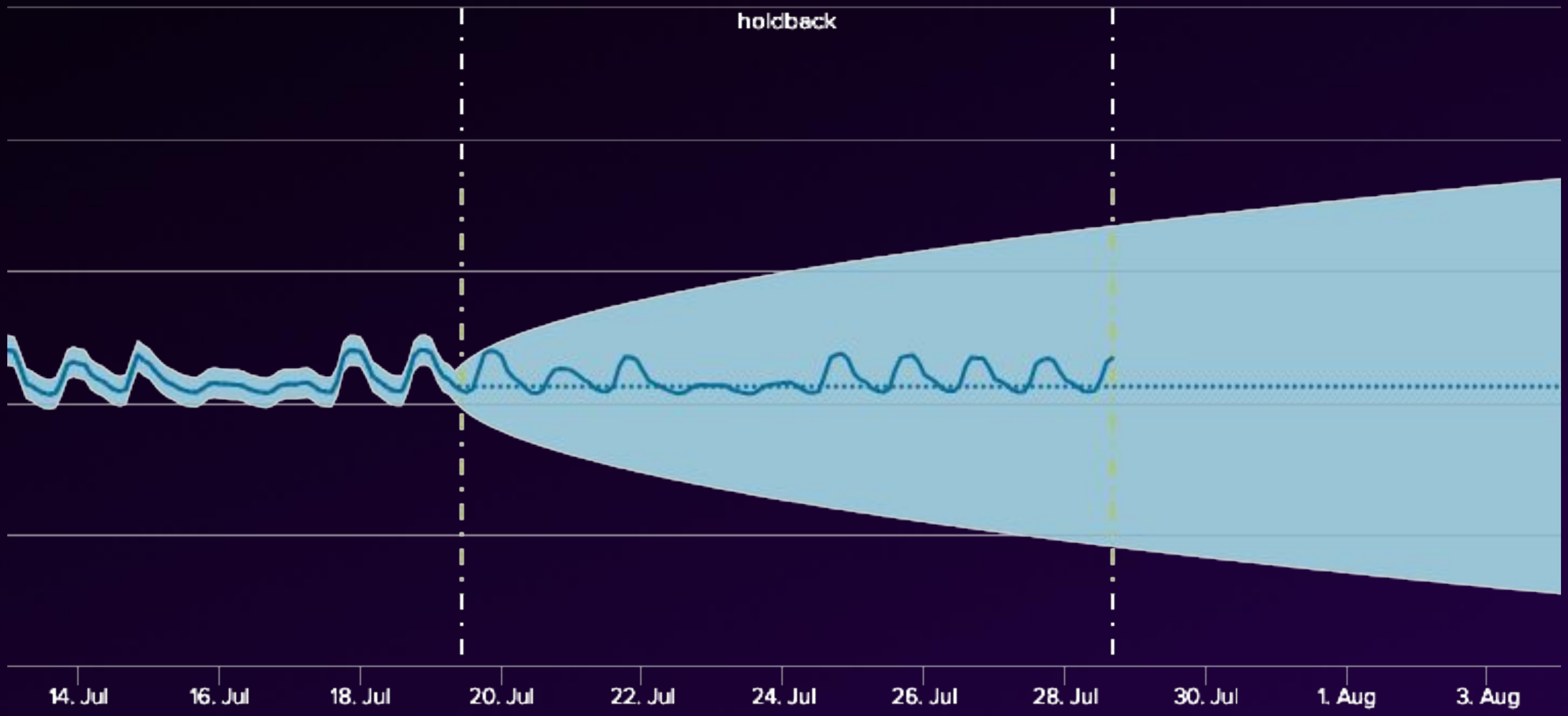
Source:Parzival’1997, flaticon.com

The Kalman Filter



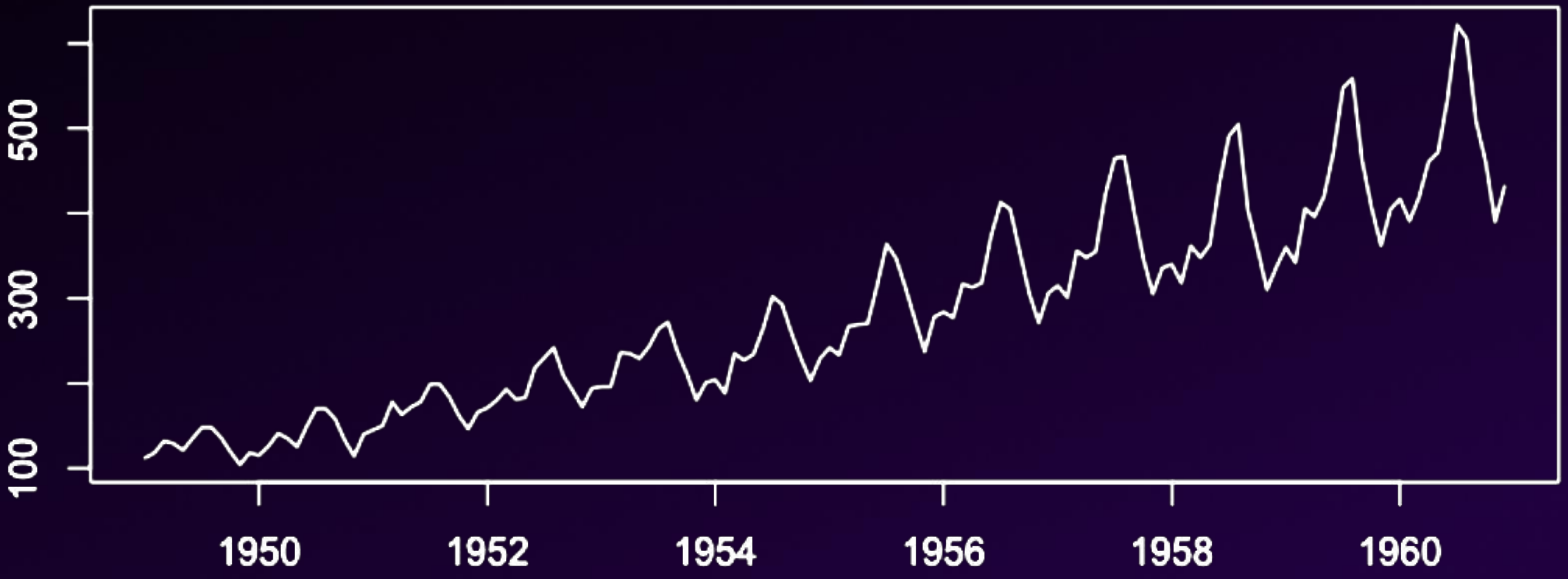
Source: https://www.kalmanfilter.net/kalman1d.html

The Kalman Filter



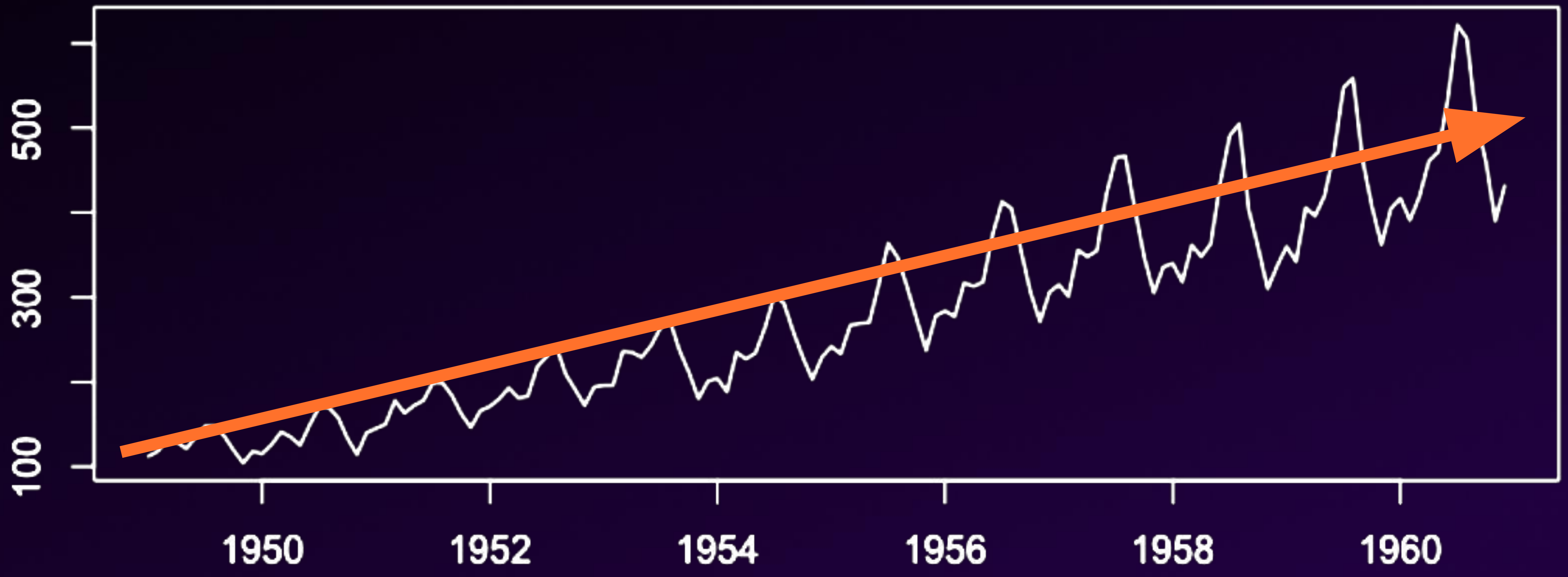
Source: https://www.kalmanfilter.net/kalman1d.html

The Kalman Filter



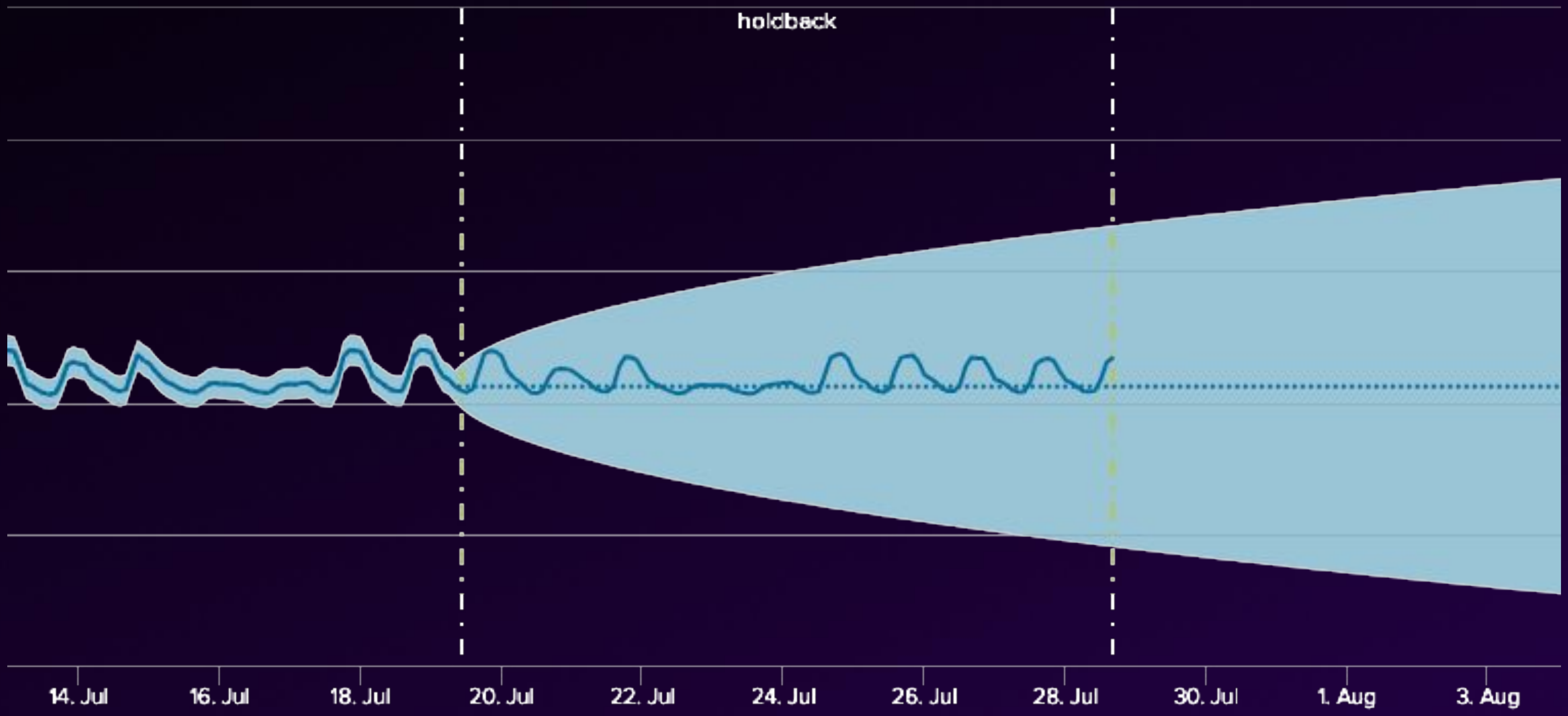
Source: https://anomaly.io/seasonal-trend-decomposition-in-r/index.html

The Kalman Filter



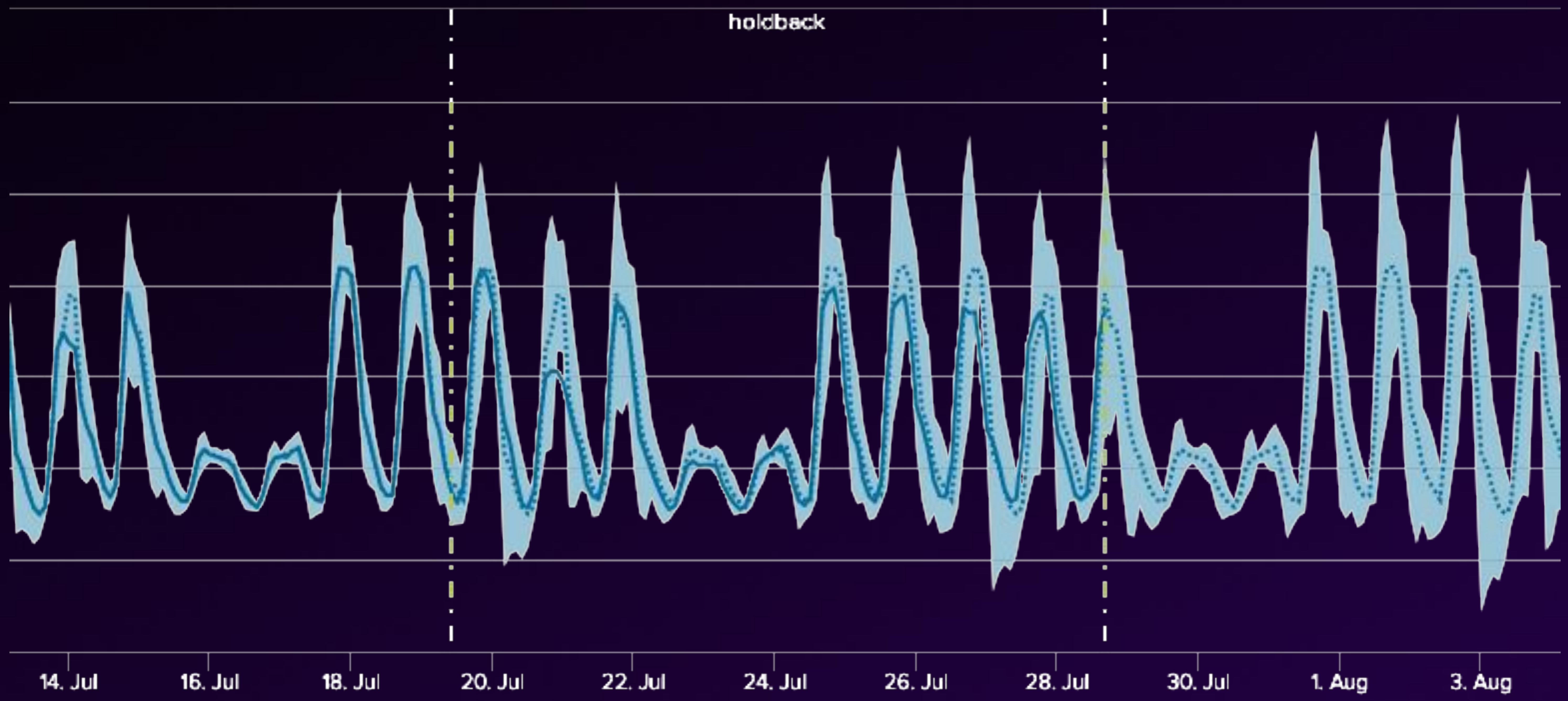
Source: https://anomaly.io/seasonal-trend-decomposition-in-r/index.html

The Kalman Filter



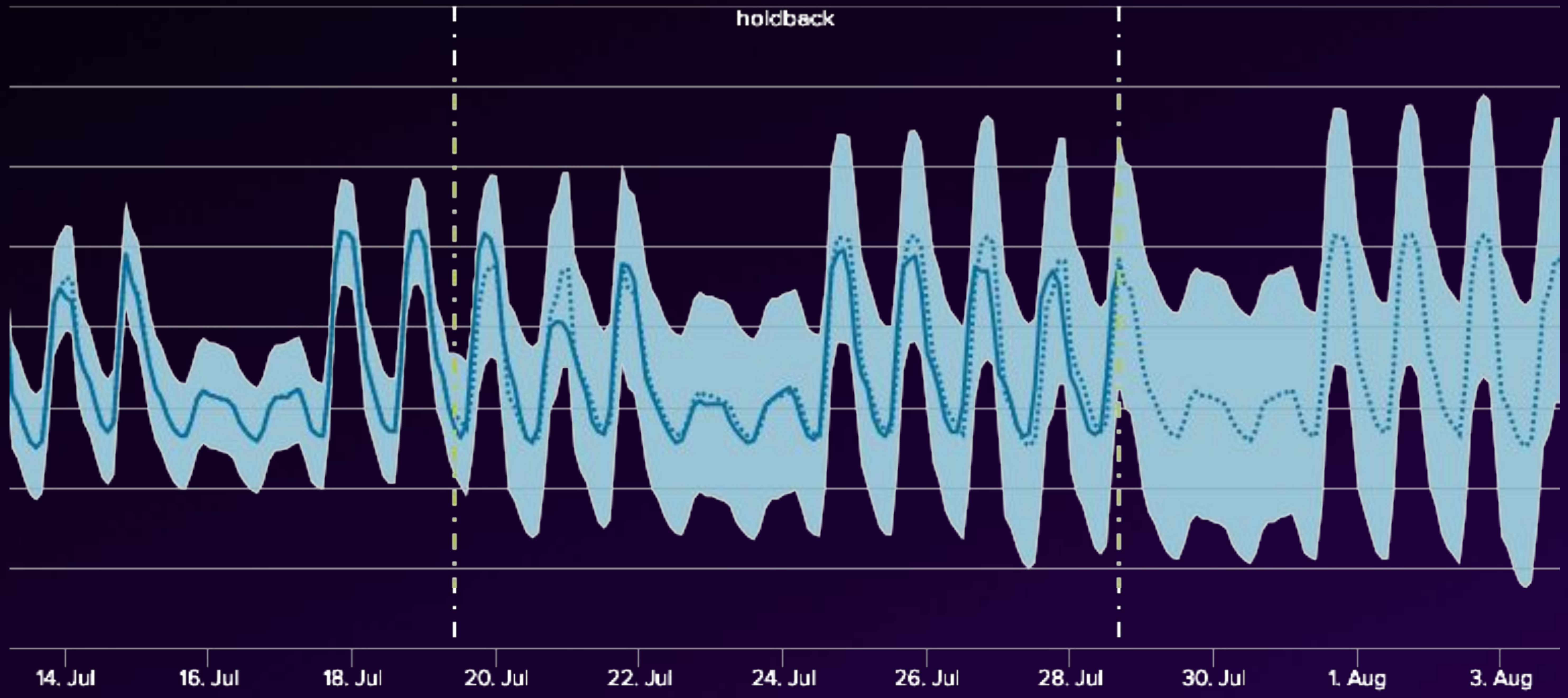
Source: https://www.kalmanfilter.net/kalman1d.html

The Kalman Filter



Source: https://www.kalmanfilter.net/kalman1d.html

The Kalman Filter



Source: https://www.kalmanfilter.net/kalman1d.html

Forecasting Time Series

"Using historical data to identify patterns, which are then used to forecast how your data might behave in the future"

Source:Parzival'1997, flaticon.com

Live Instance Demo

Log Into [INSTANCE URL]

Lab Guide Exercise #2

Time: 10 minutes

Summary

Top 4 most important things to remember about forecasting time series

1

Forecasting time series is done using a **supervised learning** method

2

Models **assume historic data as a baseline**, and will self-correct accordingly

3

Parameters have a **large impact on performance**. Tuning each model is highly recommended

4

Choice of forecasting algorithm may rely on a **subject matter expert** of the data

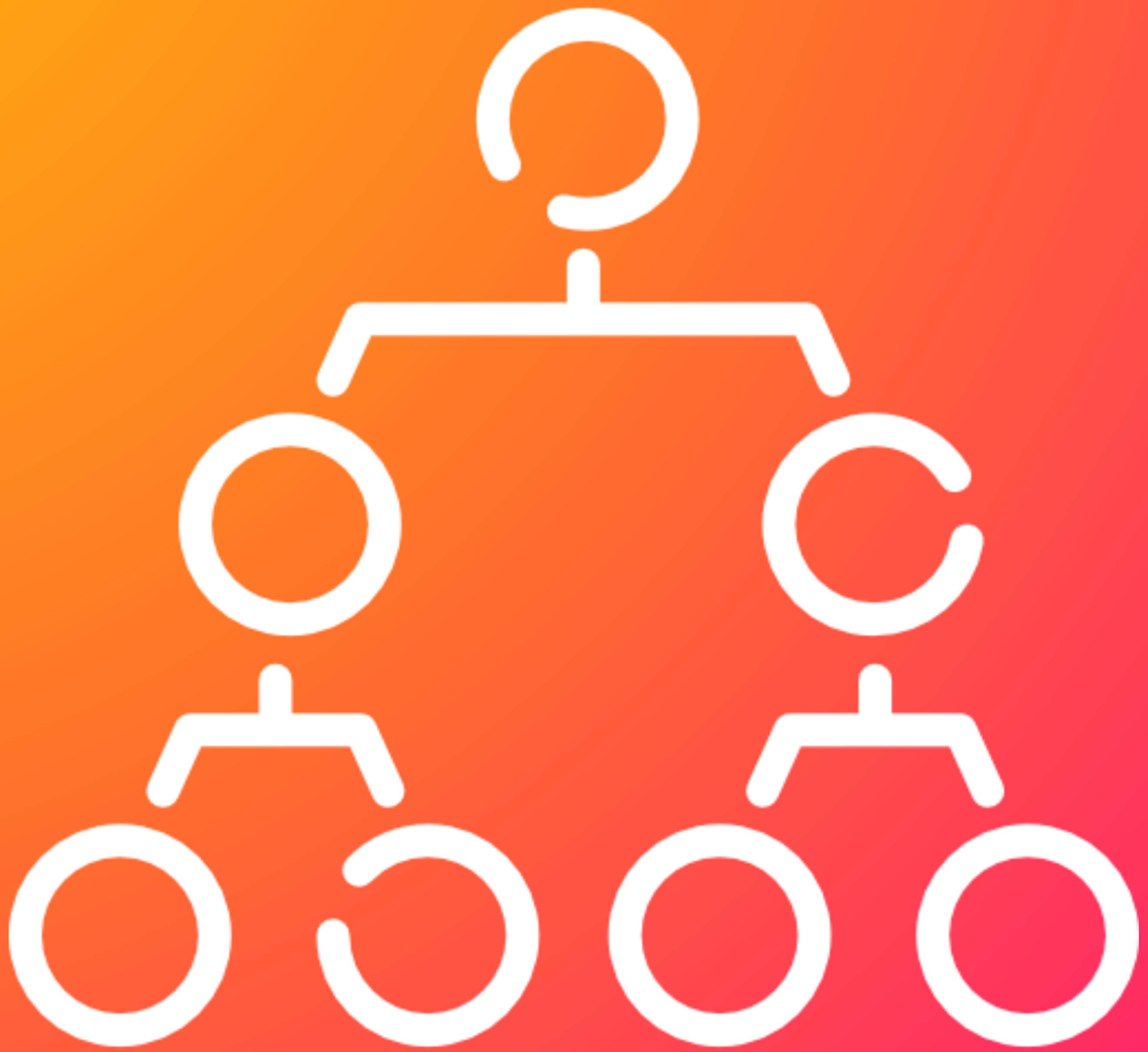
Resource Forecasting

FORECASTING END

CATEGORICAL PREDICTION START

Use Case:

Filtering out False Positive Alerts



Categorical Prediction

"A method that lets you quickly, easily, and sustainably gain insight into your data by predicting its categorical features"

How Does Categorical Prediction Work?

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCup Digital	17	btcup_checkout	375	True positive
BTCup Digital	3	payment_svc	89	True positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCup Digital	17	btcup_checkout	375	True positive
BTCup Digital	3	payment_svc	89	True positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCup Digital	17	btcup_checkout	375	True positive
BTCup Digital	3	payment_svc	89	True positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCup Digital	17	btcup_checkout	375	True positive
BTCup Digital	3	payment_svc	89	True positive

alert group	count	src	alert value	Predicate (predicted)
T-Shirt Co.	17	checkoutsvc	100	False positive
O11y Cloud	16	browsercheck	89	False positive

How Does Categorical Prediction Work?

alert group	count	src	alert value	Predicate
T-Shirt Co.	15	checkoutsvc	107	False positive
BTCup Digital	17	btcup_checkout	375	True positive
BTCup Digital	3	payment_svc	89	True positive

alert group	count	src	alert value	Predicate (predicted)
T-Shirt Co.	17	checkoutsvc	100	False positive
O11y Cloud	16	browsercheck	89	True positive?

How Does Categorical Prediction Work?

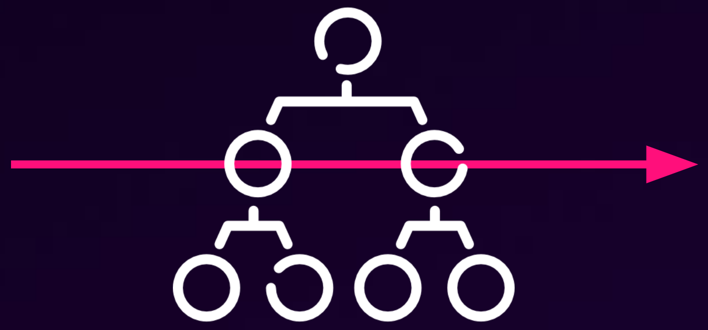
Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL

How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL



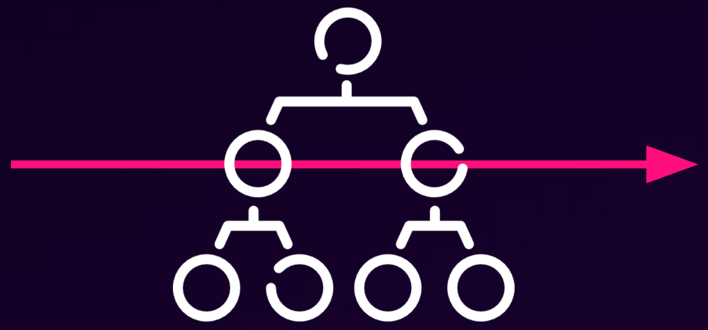
New Data

Feature 1	...	Feature n	Severity (predicted)
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL

How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL



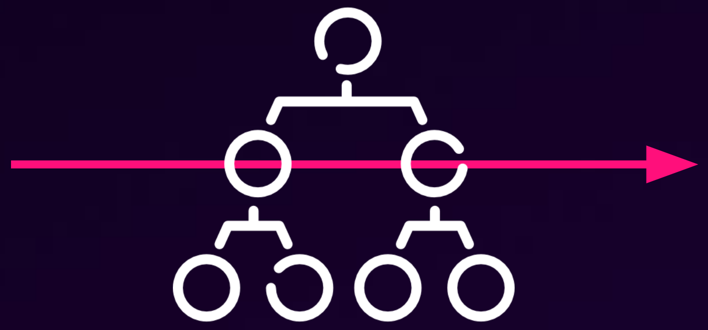
New Data

Feature 1	...	Feature n	Severity (predicted)	Severity (actual)
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[data]	...	[data]	CRITICAL	CRITICAL
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	MINOR

How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL



New Data

Feature 1	...	Feature n	Severity (predicted)	Severity (actual)
[data]	...	[data]	CRITICAL	MAJOR
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[data]	...	[data]	CRITICAL	CRITICAL
[data]	...	[data]	CRITICAL	LOW
[data]	...	[data]	CRITICAL	MINOR
[illegible]	...	[data]	CRITICAL	MAJOR
[illegible]	...	[data]	CRITICAL	MAJOR
[illegible]	...	[data]	CRITICAL	MINOR

How Does Categorical Prediction Work?

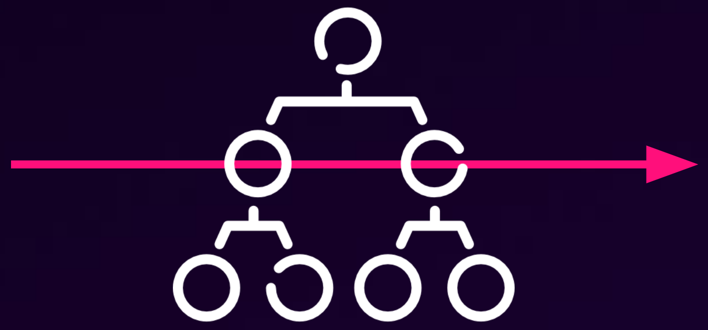
Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	MAJOR
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR

How Does Categorical Prediction Work?

Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	MAJOR
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR



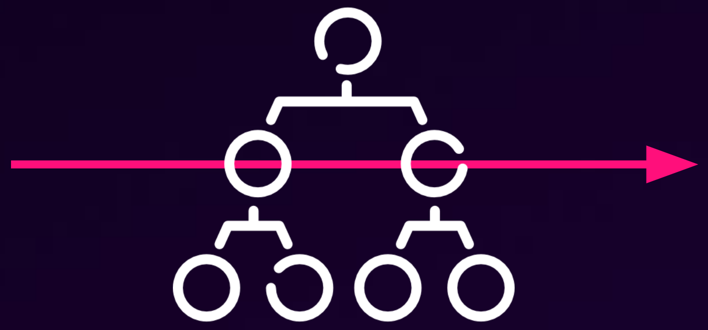
New Data

Feature 1	...	Feature n	Severity (predicted)
[data]	...	[data]	LOW
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	LOW

How Does Categorical Prediction Work?

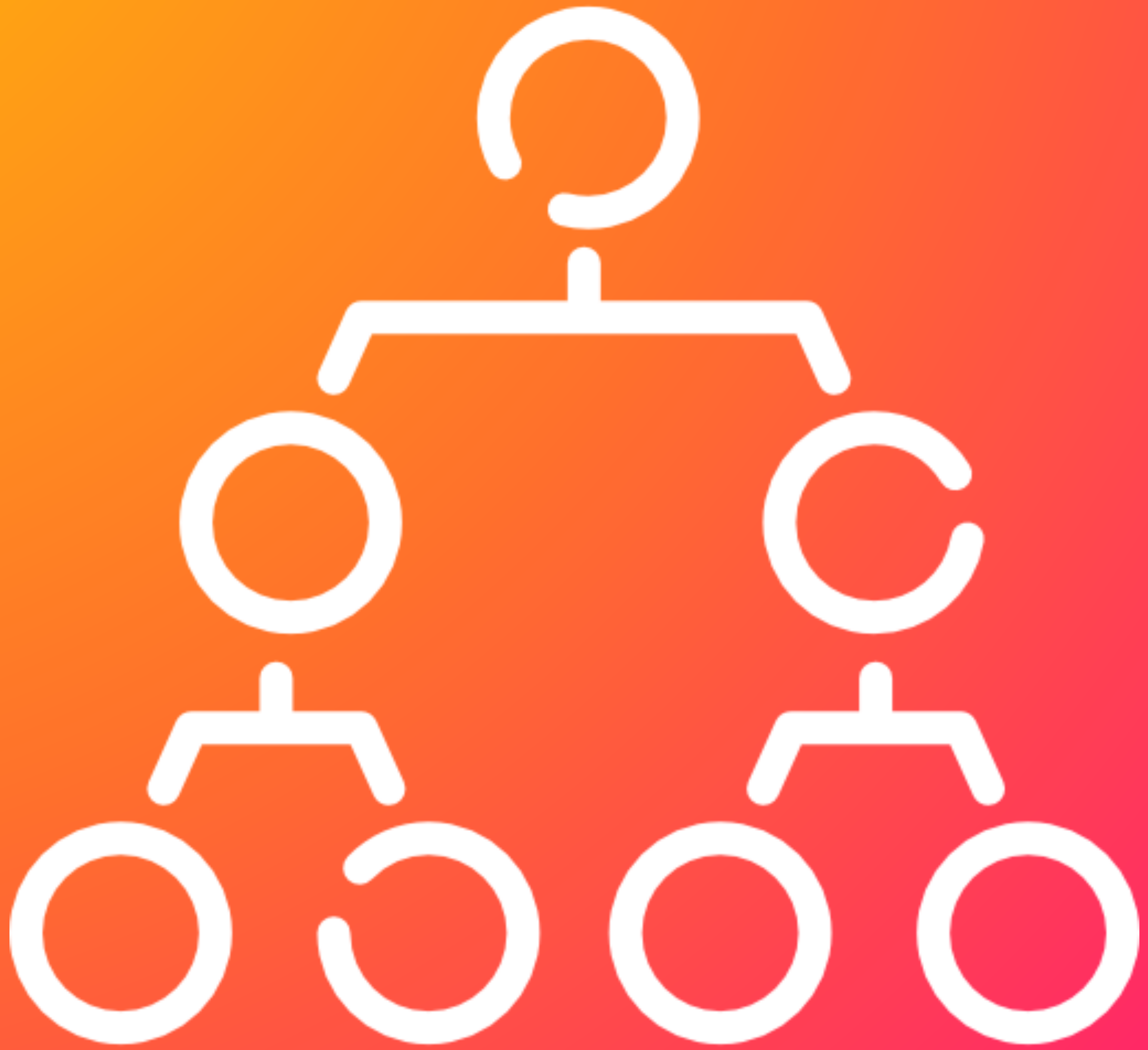
Training Data

Feature 1	...	Feature n	Severity
[data]	...	[data]	MAJOR
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	LOW
[data]	...	[data]	MINOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR



New Data

Feature 1	...	Feature n	Severity (predicted)
[data]	...	[data]	LOW
[data]	...	[data]	MAJOR
[data]	...	[data]	MAJOR
[data]	...	[data]	MINOR
[data]	...	[data]	CRITICAL
[data]	...	[data]	CRITICAL
[illegible]	...	[data]	LOW
[illegible]	...	[data]	MINOR
[illegible]	...	[data]	LOW



Categorical Prediction

"A method that lets you quickly, easily, and sustainably gain insight into your data by predicting its categorical features"

Live Instance Demo

Log Into [INSTANCE URL]

Lab Guide Exercise #5
Time: 10 minutes

Summary

Top 4 most important things to remember about categorical prediction

1

Predicting categorical fields is done using a **supervised learning** method which uses **labeled data**

2

Models **assumes there exists a pattern** determining existing categories

3

Scaling data prior to training is **often necessary**

4

Choice of categorical prediction algorithm(s) may rely on a **subject matter expert** of the data

CATEGORICAL PREDICTION END

OUTLIER DETECTION START

Use Case:

Detecting Outliers in CPU Utilization

Outlier Detection Algorithms

Identify and analyze abnormal behavior in your data

Includes:

- Clustering
- Outlier Detection

Source:Hajicon, flaticon.com

Global

Data points different from expected pattern, range, or norm

Contextual

Are the results out of context?

Collective

Looks normal with isolation but stands out in a group

Numeric

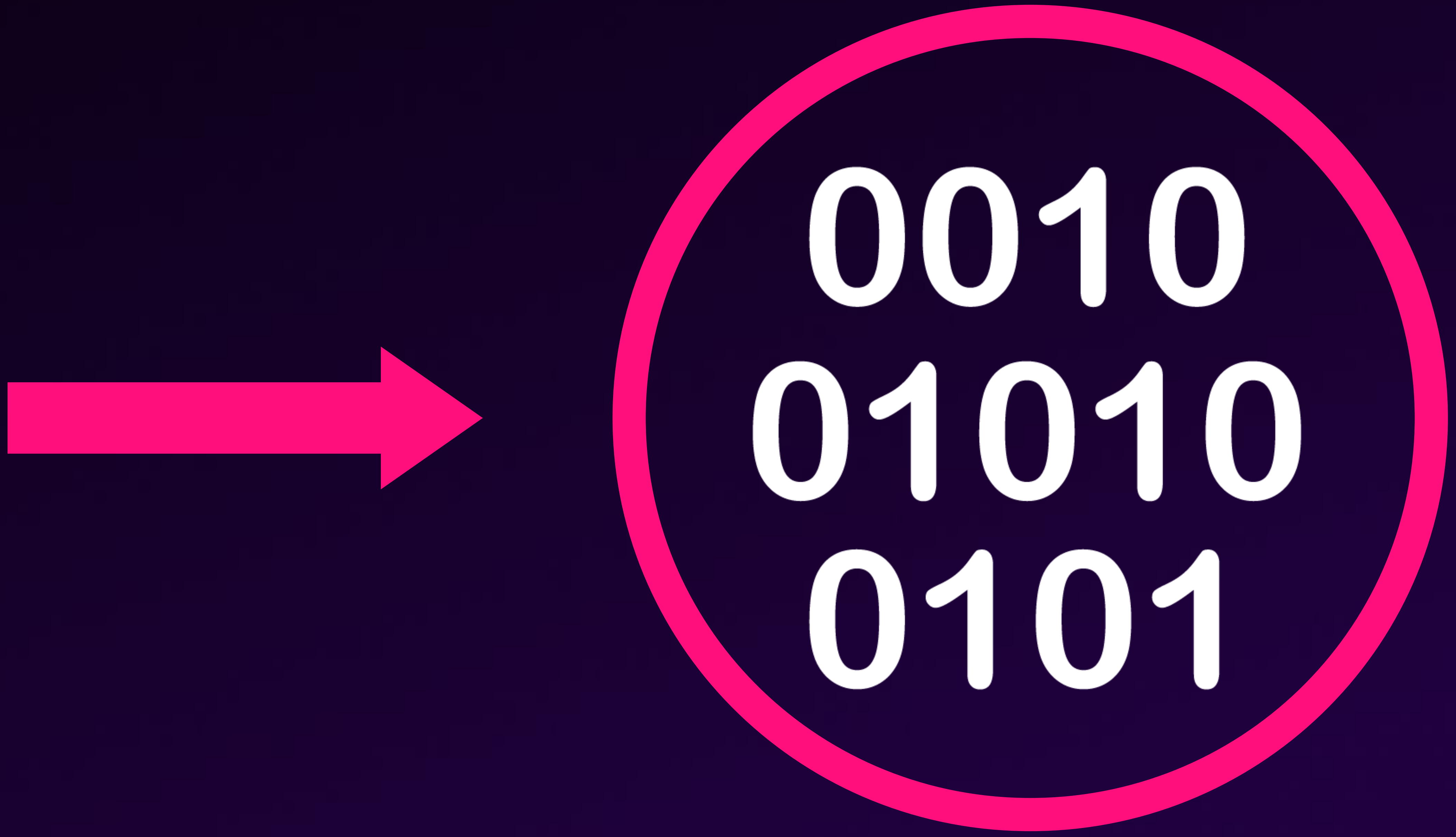
Categorical

Included Algorithms

DensityFunction | LocalOutlierFactor | MultiVariateOutlierDetection | One-Class SVM

Categorical data to Numeric data?

device
server01
server02
server03



Categorical data to Numeric data?

device
server01
server02
server03

device	server01	server02	server03
server01	1	0	0
server02	0	1	0
server03	0	0	1

Outlier Detection Algorithms

Identify and analyze abnormal behavior in your data

Includes:

- Clustering
- Outlier Detection

Source:Hajicon, flaticon.com

Live Instance Demo

Log Into [INSTANCE URL]

Lab Guide Exercise #3
Time: 10 minutes

Summary

Top 4 most important things to remember about outlier detection

1

Outlier detection is a way of analyzing your data for **historical baseline outliers**

2

Models **assume historic data input represents normal data**

3

Encoding is necessary for categorical outlier detection

4

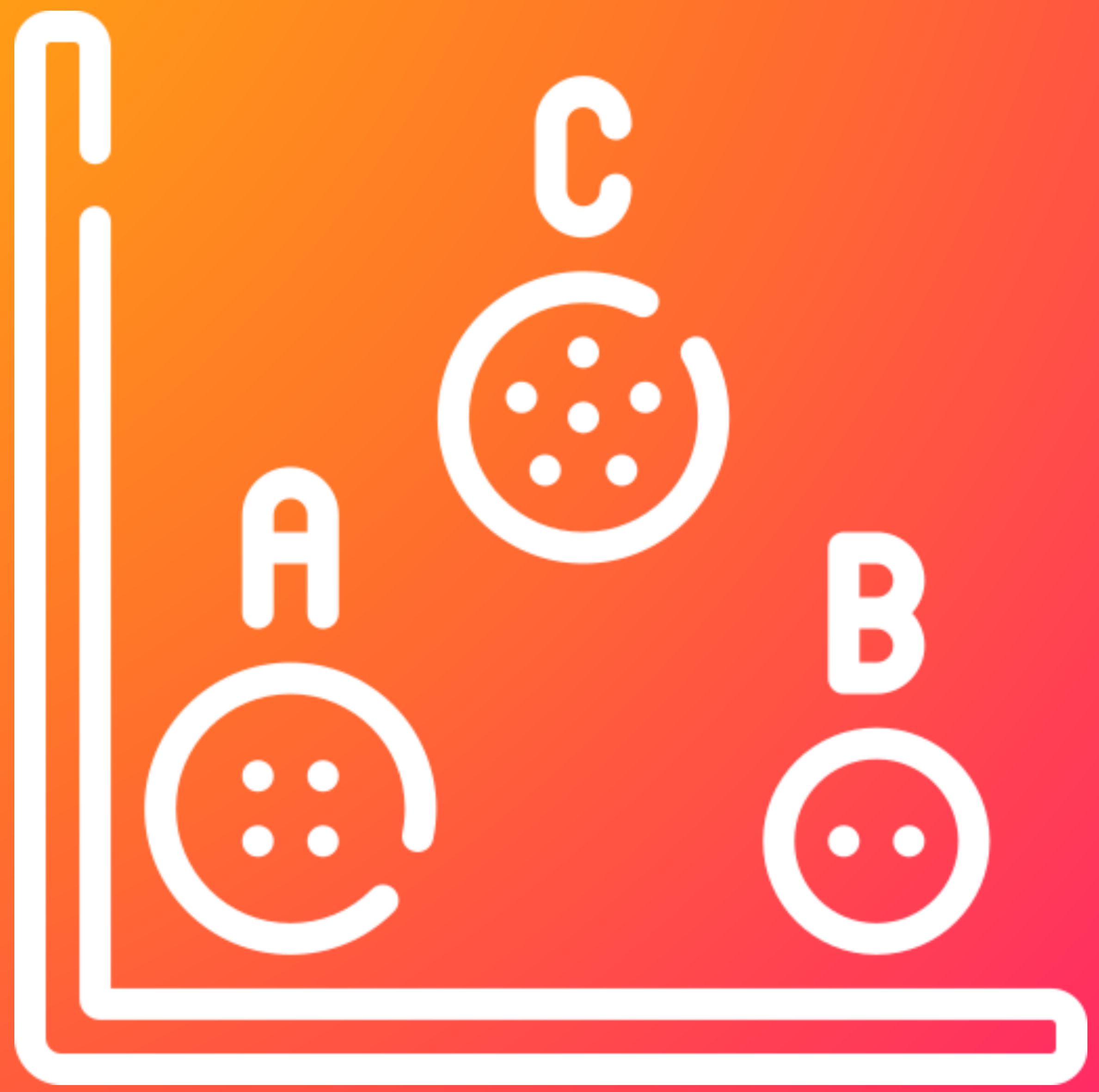
Choice of outlier algorithm may rely on a **subject matter expert** of the data

Adaptive Thresholding

OUTLIER DETECTION END

CLUSTERING START

Use Case:
Referencing Historically Similar Tickets



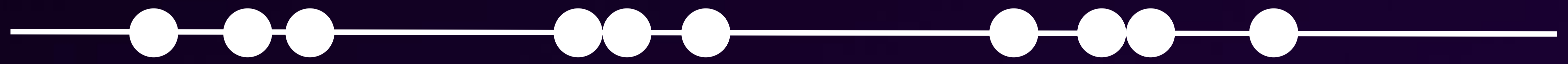
Clustering

"A method that organizes a set of numeric data points in a way that objects in the same cluster are more similar to each other than those in other clusters"

How Does Numeric Clustering Work?

General intuition

Starting with one-dimensional data, unlabeled numeric values plotted on a line



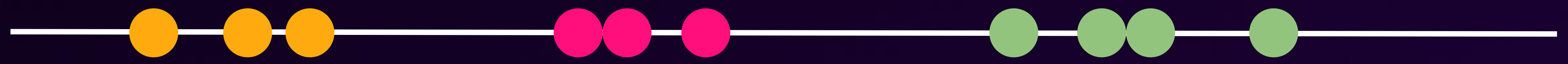
Organizing Exercise: Using your human intuition, form 3 groups with the dataset above

Source: StatQuest: K-means clustering

How Does Numeric Clustering Work?

General intuition

Most likely, these are the 3 groups (or clusters) you formed:



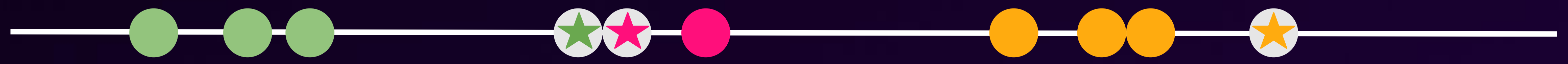
Now let's see how a machine can replicate this kind of grouping intuition!

Source: StatQuest: K-means clustering

How Does Numeric Clustering Work?

General intuition

First, estimate the center of a cluster randomly, denoted by the star



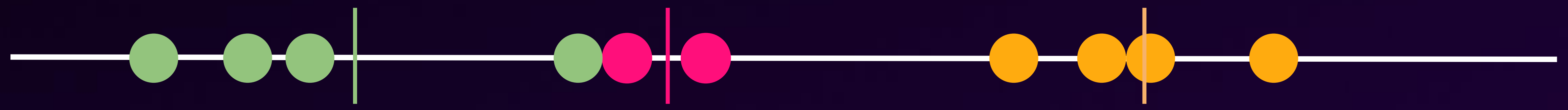
So, our clusters look like this:

Source: StatQuest: K-means clustering

How Does Numeric Clustering Work?

General intuition

Now the **mean** value of each cluster is calculated. This is the “mean” in “k-means”!

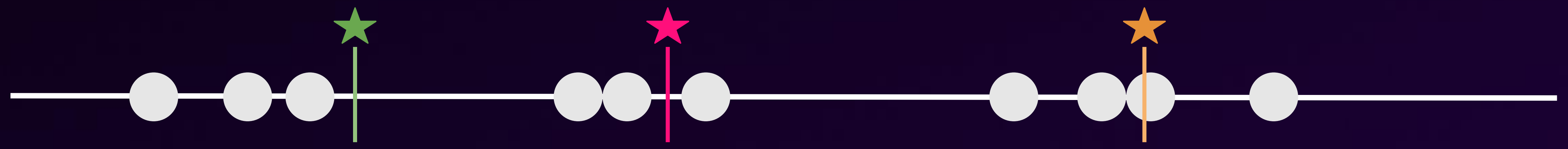


Source: StatQuest: K-means clustering

How Does Numeric Clustering Work?

General intuition

The mean of each cluster now become the new centroids:

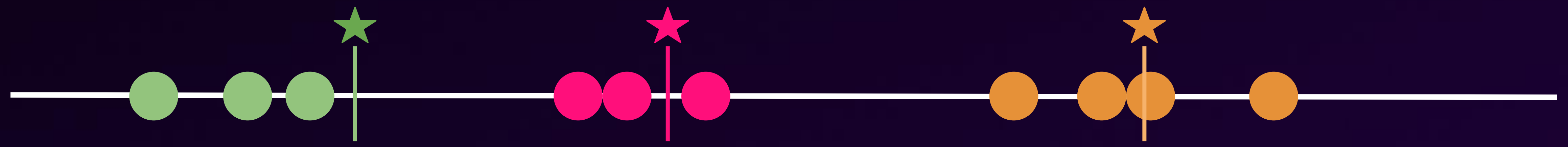


Source: StatQuest: K-means clustering

How Does Numeric Clustering Work?

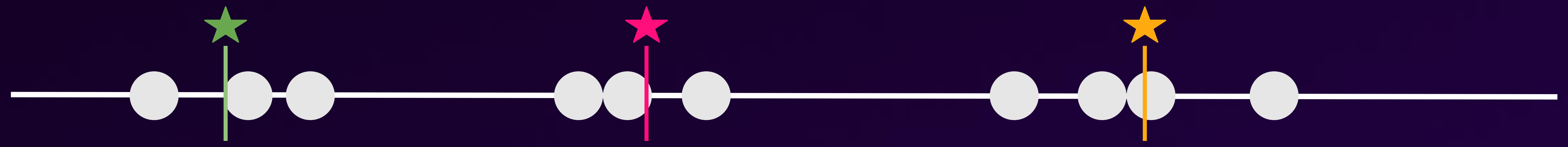
General intuition

Now the data points are reassigned to the nearest k-means cluster:

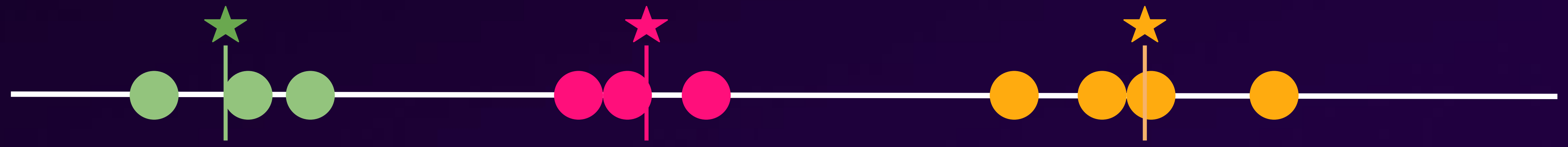


The k-means and clustering steps are repeated until the data points no longer change to different clusters

Find the cluster mean:



Reassign datapoints:

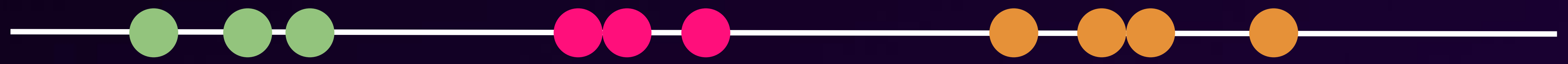


Source: StatQuest: K-means clustering

How Does Numeric Clustering Work?

General intuition

These are our final 3 clusters, found by k-means clustering:

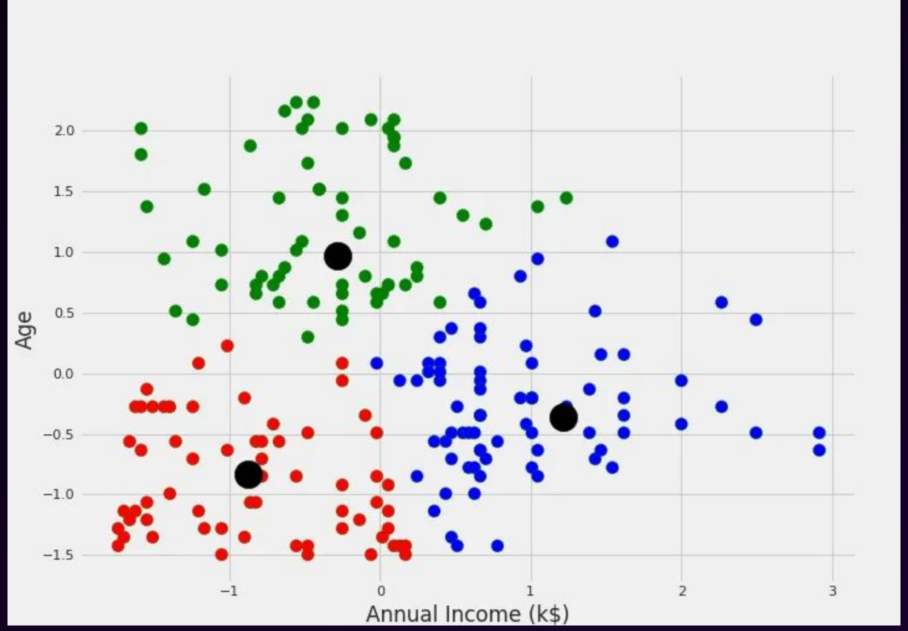


Source: StatQuest: K-means clustering

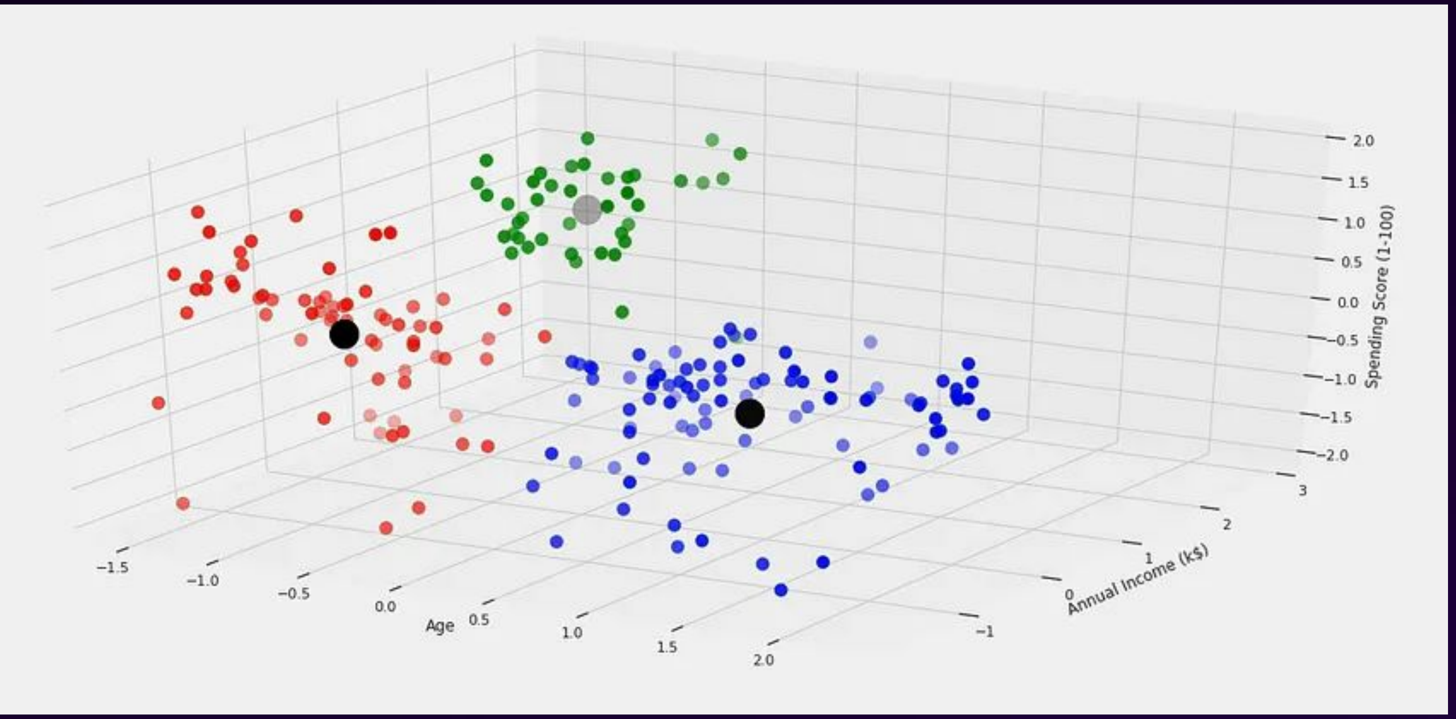
How Does Numeric Clustering Work?

General intuition

k-means clustering applies to multidimensional data as well:



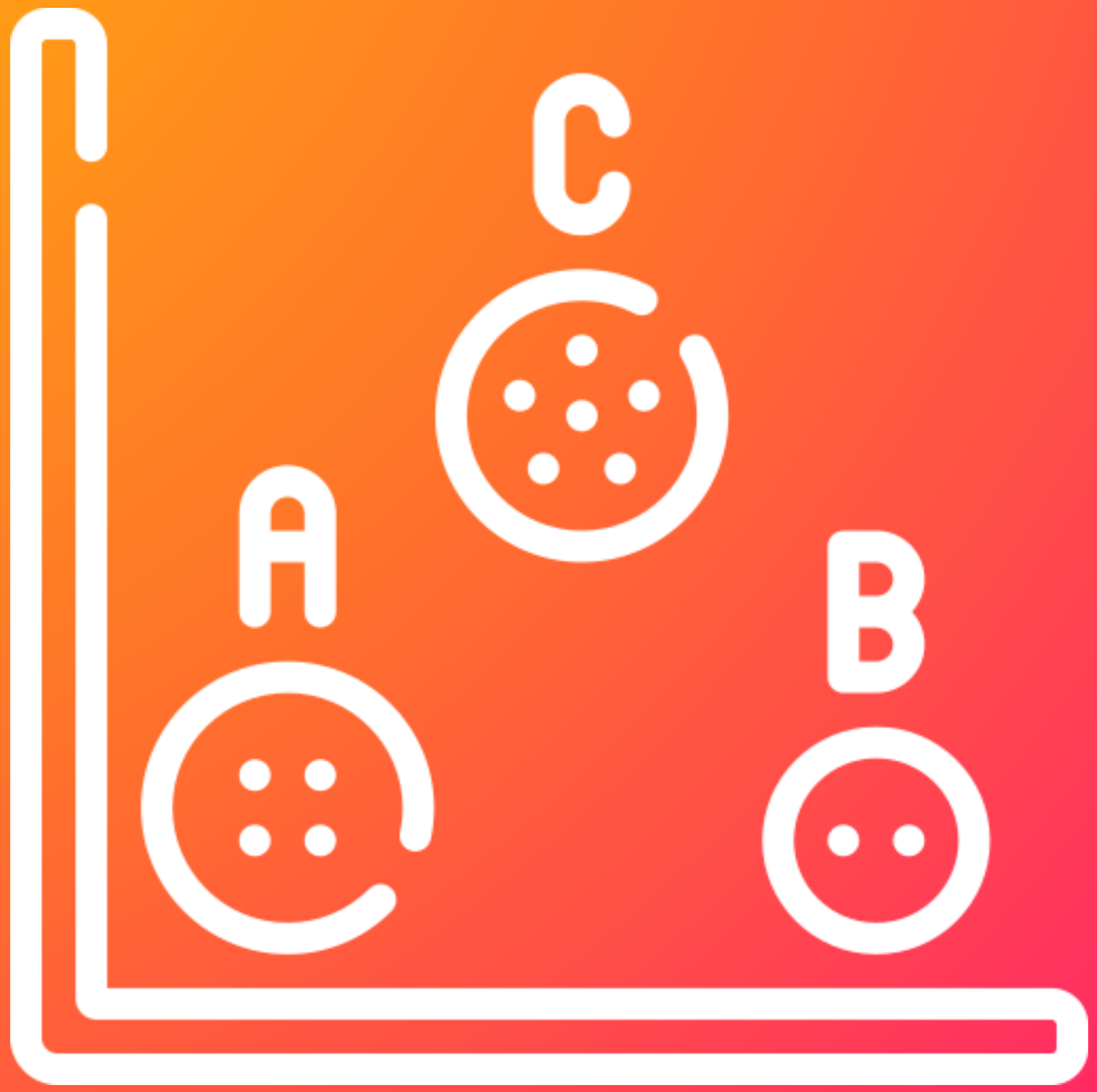
2 dimensional data



3 dimensional data

If you can see in more than 3 dimensions… give us a call

4+ dimensional data



Clustering

"A method that organizes a set of numeric data points in a way that objects in the same cluster are more similar to each other than those in other clusters"

Live Instance Demo

Log Into [INSTANCE URL]

Lab Guide Exercise #4

Time: 10 minutes

Summary

Top 4 most important things to remember about clustering algorithms

1

Clustering is an **unsupervised learning** method which uses **unlabeled data**

2

User must decide whether **dimensionality reduction** is necessary

3

User must decide whether to **scale** the data prior to clustering

4

Choice of clustering algorithm(s) may rely on a **subject matter expert** of the data

Alert Storm Detection

CLUSTERING END

And we’re done!

Summary

Tools in Your ML Toolkit Now

Prediction

Get ahead of issues that may happen in the future

Categorization

Uncover insights about your data to quickly respond in the present

Outlier Detection

Identify and analyze abnormal behavior in your data

Attribution: Freepik and Becris from flaticon.com

Additional Resources

Getting started

- View some of our webinars
- Check out our YouTube playlist
- Check out the blog on MLTK 5.4 release
- Try out some of our starter blogs, such as Cyclical Statistical Forecasts and Anomalies, part 1
- Try our new MLTK Deep Dives

Increasing complexity

- Try part 4 or 6 of the Cyclical Statistical Forecasts and Anomalies series
- Brush up on how MLTK works with our comprehensive documentation
- Get familiar with the Workshop Guide

More advanced

- The Analytics and Data Science course
- Try out the Anomalies Are Like a Gallon of Neapolitan Ice Cream - Part 1
- Try out part 5 of the cyclical statistical forecasts and anomalies series
- Try the ML-SPL API

Thank you!

.conf24
splunk>

Q&A

.conf24
splunk>