

How to Think About Machine Learning Algorithms

INTRODUCING MACHINE LEARNING



Swetha Kolalapudi

CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Spot applications of Machine Learning in the real world

Differentiate between the different types of Machine Learning problems

Pick your problem: Classify, Regress, Recommend or Cluster

Machine Learning

is the invisible hand
behind so many things we
like to take for granted



Services we love to use...



Netflix



Amazon

Services we love to use...

Netflix

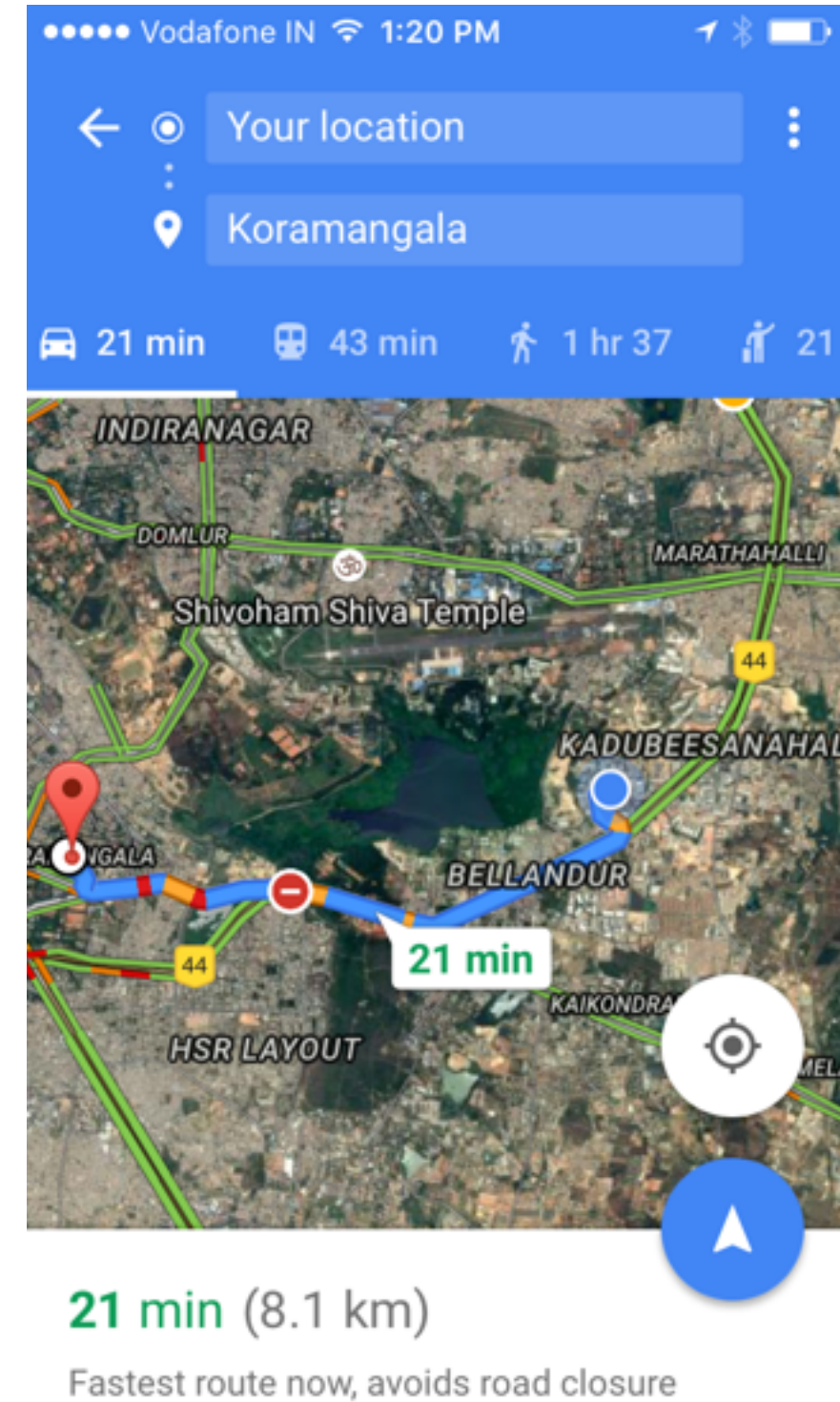
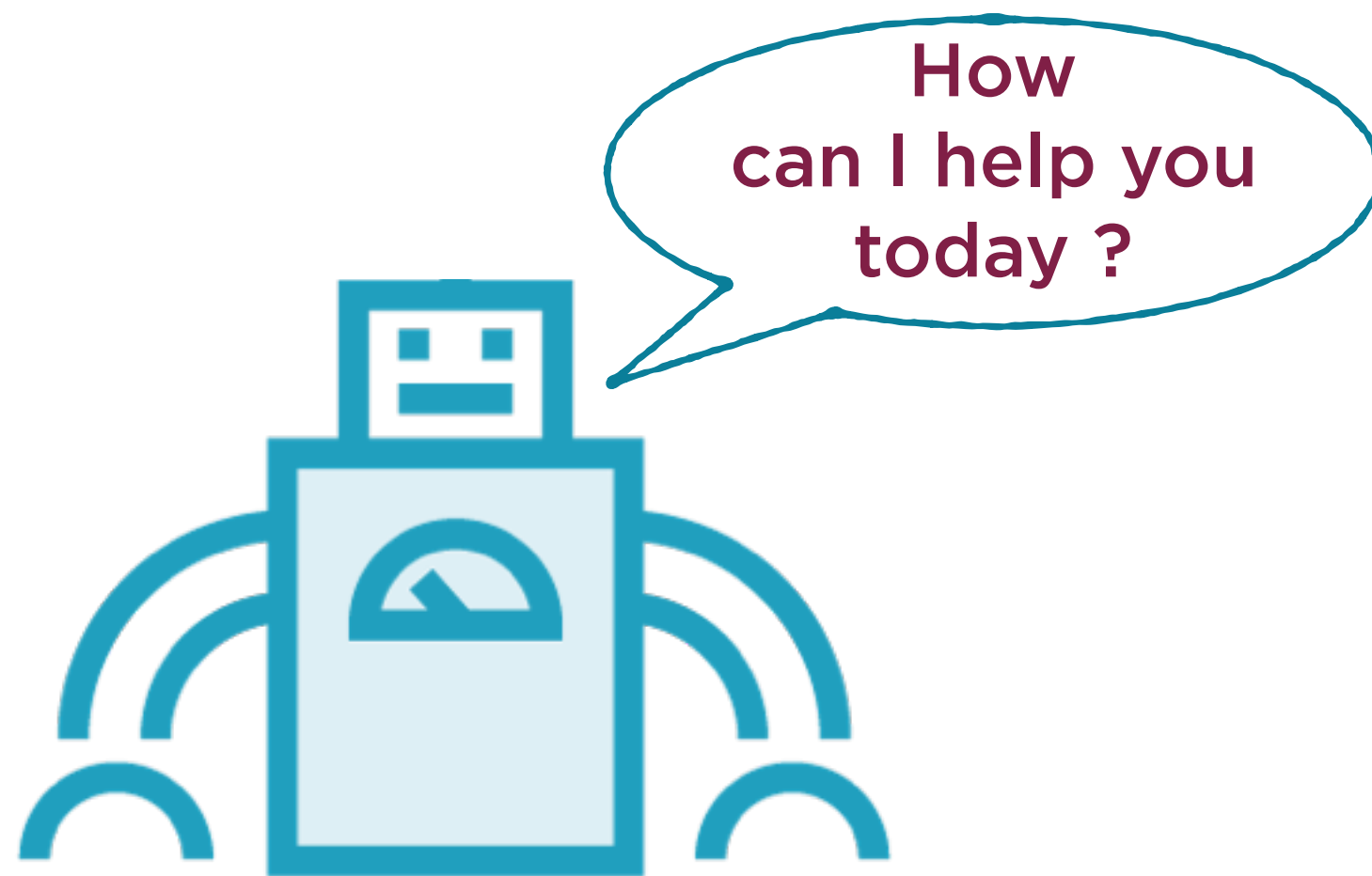
Amazon

... because they seem to know
exactly what we'll like or need

Recommended
for you!



Applications that are making themselves indispensable to our daily routine



The little details that make every task that much easier and faster

Inbox (11,764)

Starred

Important

Chats

All Mail

Spam (577)

autoc|omplete



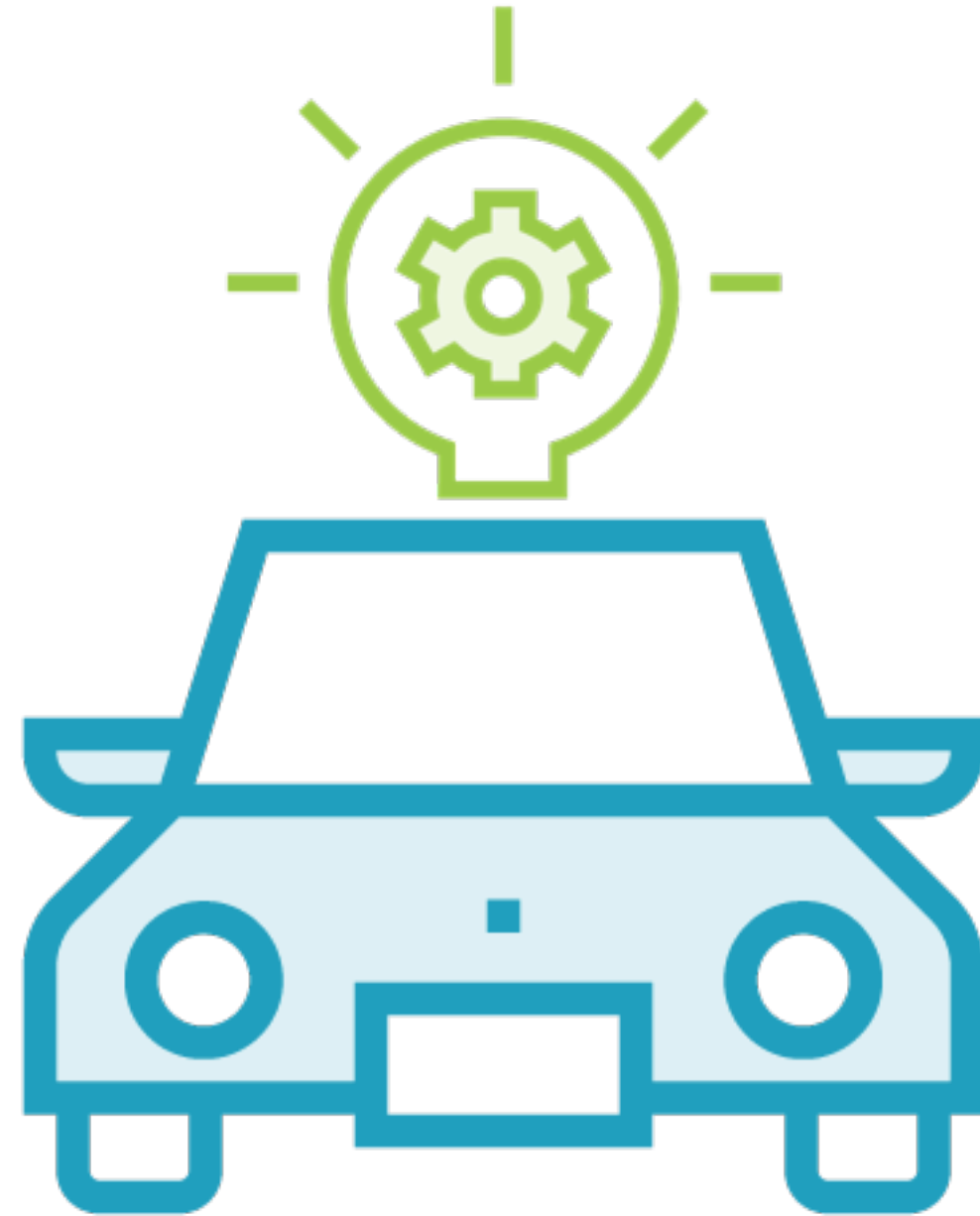
autocomplete

autocomplete api

autocomplete algorithm

autocomplete address

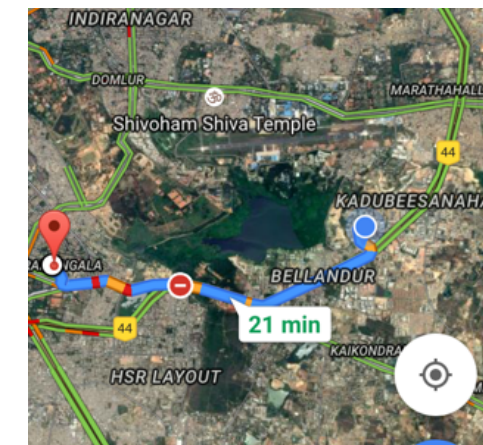
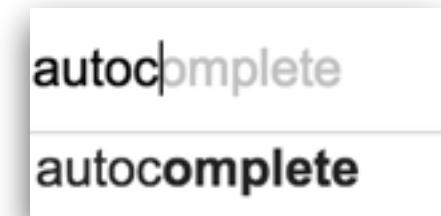
Innovations that are
ushering in the future



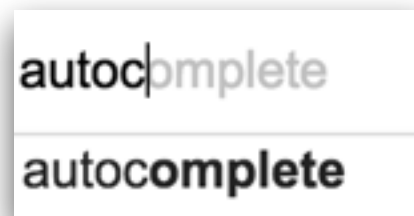
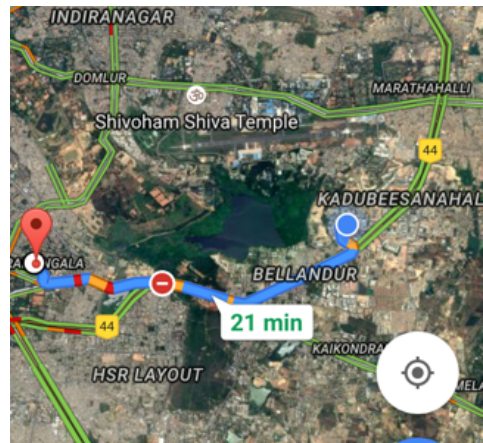
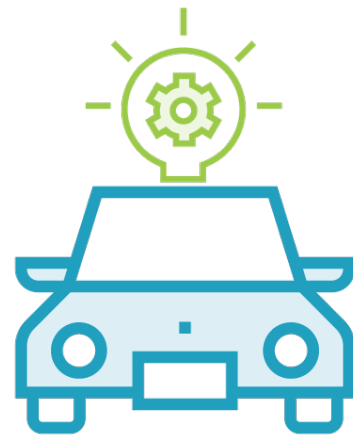
Inbox (11,764)



Machine Learning is the invisible hand behind all of these



Inbox (11,764)



What is Machine Learning?

What makes it so perfect for such a wide variety of applications?

Let's take an example to help us understand..



Say you're an alien
observing a park on Earth



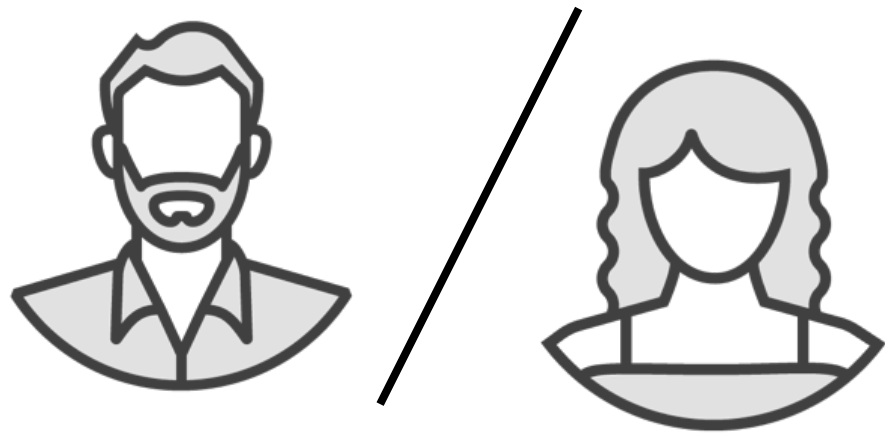
And you want to classify visitors
to the park as



Male



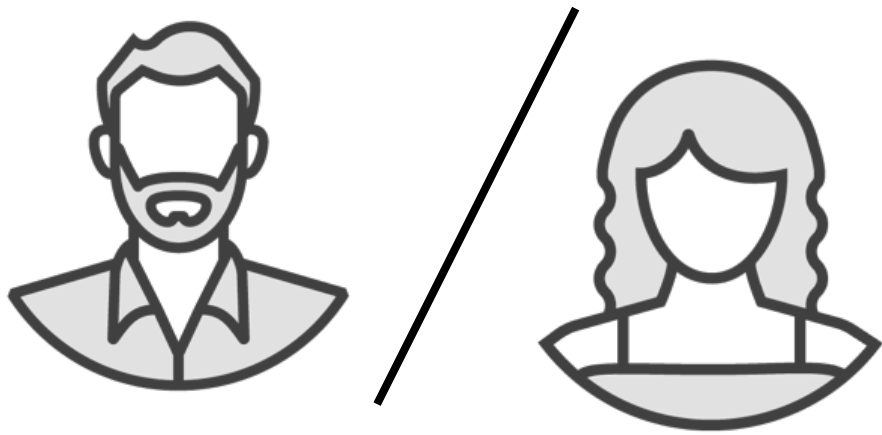
Female



If height > 6 ft, Male

Else, Female

Rule Based
Approach



1. A human guide pointing out the gender
2. Intuitively learn how to differentiate

Machine Learning
Based Approach

Machine Learning Based Approach



**The ML based
approach is similar to
how humans learn**

How Humans Learn

Human beings learn to identify patterns when they're exposed to a phenomenon for a prolonged period of time

Human beings learn from
“Experience”

Machine Learning

**A computer program/system
that can learn from
“Experience”**

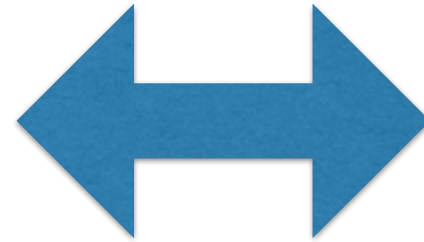
Machine Learning



As more and more data is captured, Machine learning is finding a wide variety of uses

Recommendation Systems

“Experience”



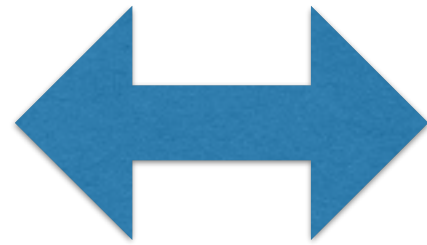
**User Clicks/
Views**

**Recommended
for you!**



Siri

“Experience”

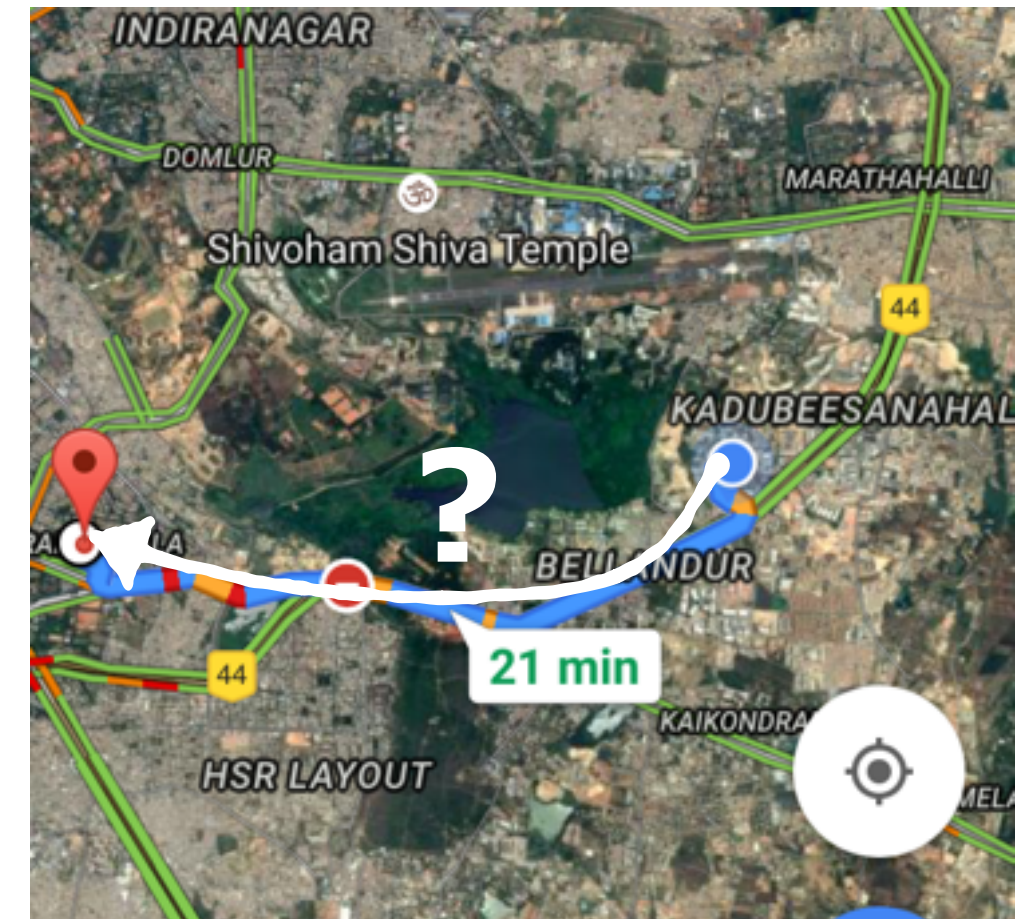


Past user
questions and
answers



Commute Time Calculation

How long will it
take to travel from
point A to **point B**?



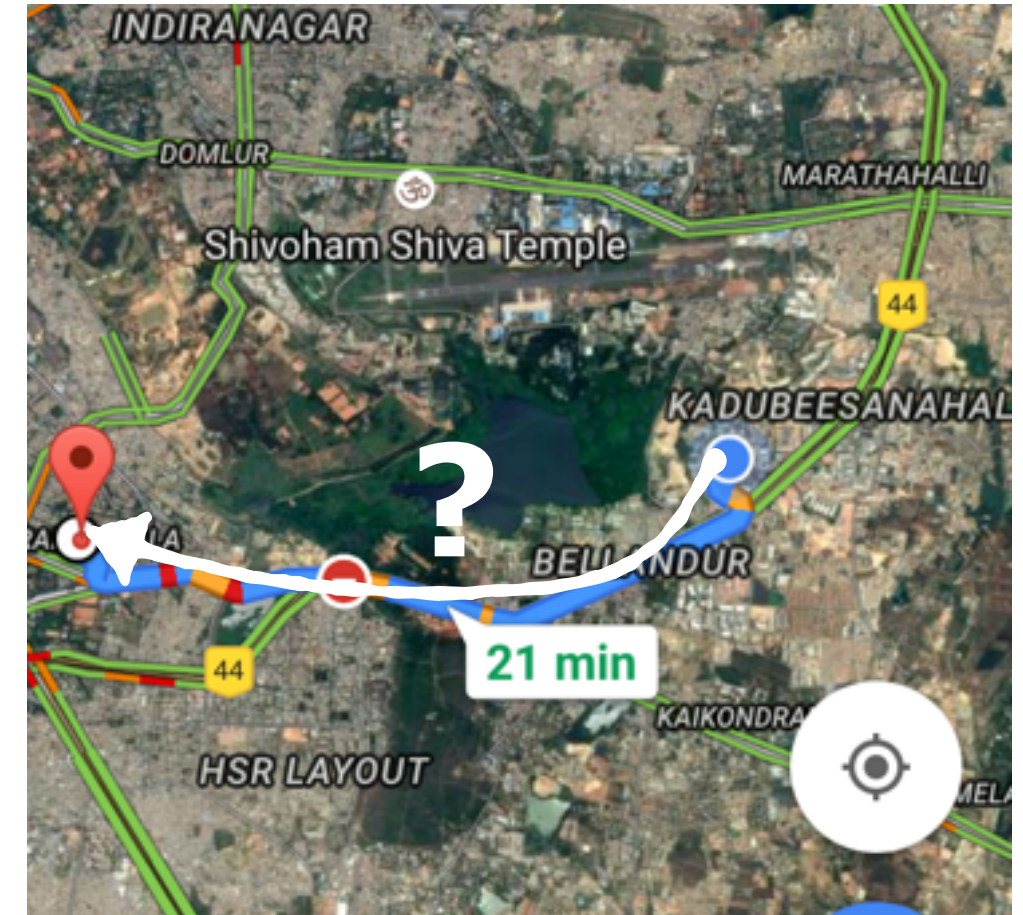
Two Approaches

Rule Based

**Machine
Learning Based**

Rule Based Approach

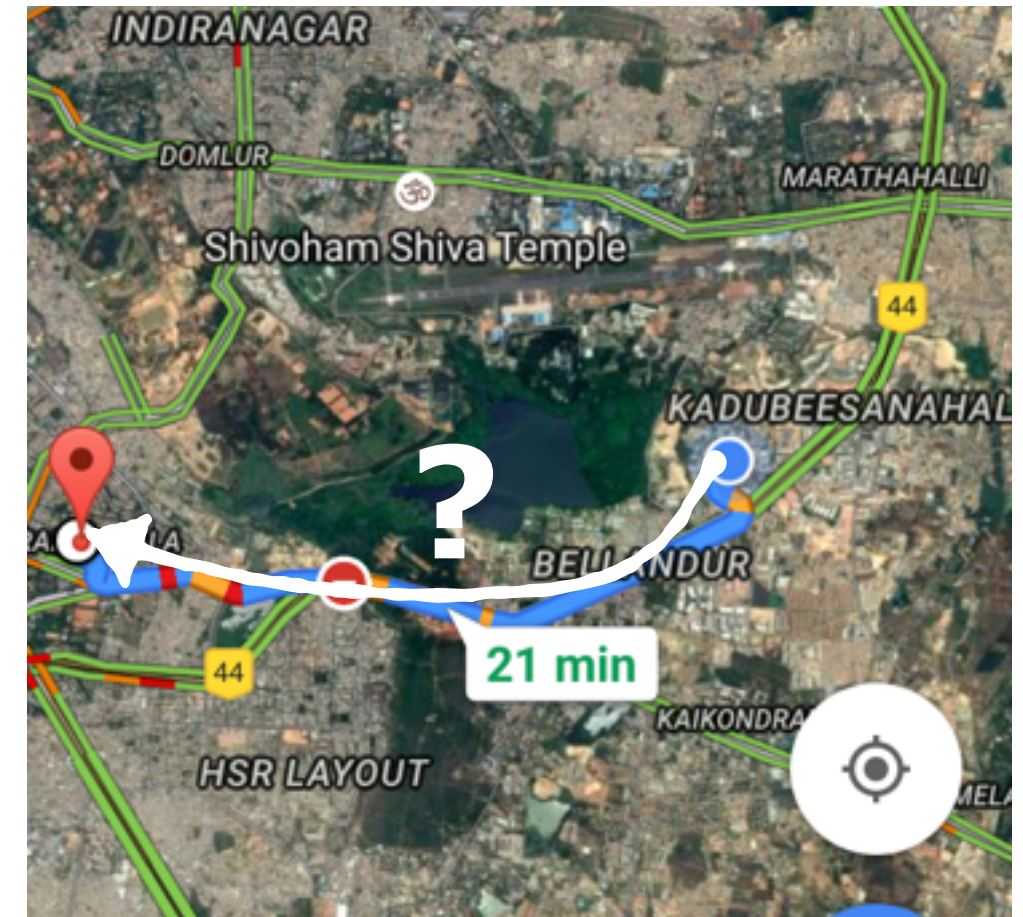
A large set of rules defined by Humans



Rule Based Approach

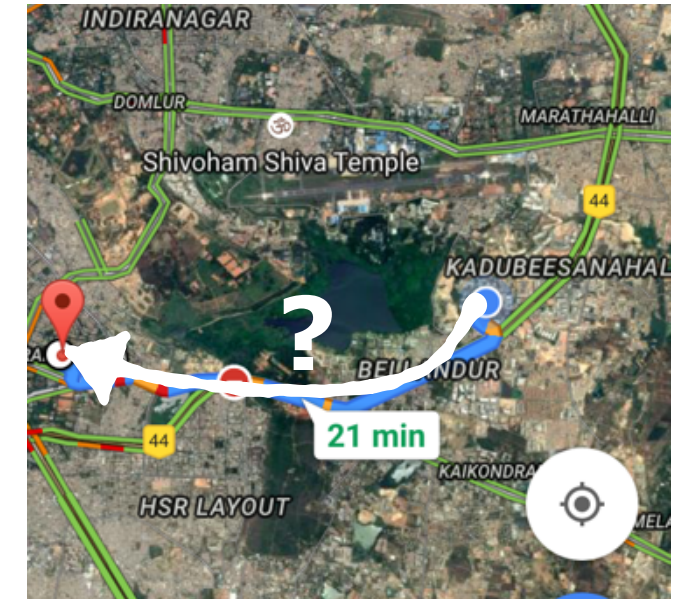
If & & Then

Day of week	Time of Day	Distance	Commute Time in Mins
Monday	10 AM - 11 AM	2 - 4 KM	40



Rule Based Approach

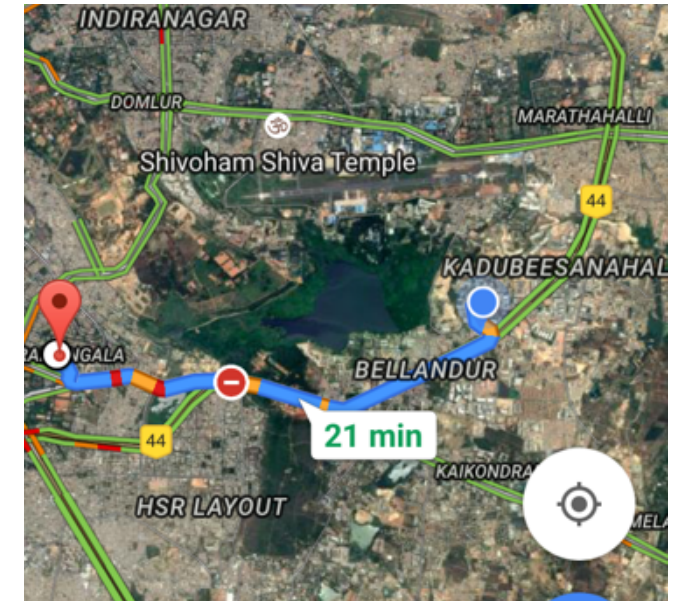
Day of week	Time of Day	Distance	Commute Time in Mins
Monday	10 AM - 11 AM	2 - 4 KM	40
Sunday	9 AM - 10 AM	4 - 6 KM	20
Tuesday	5 PM - 6 PM	2 - 4 KM	30
Friday	6 PM - 7 PM	4 - 6 KM	60



These rules are identified manually after a lot of research

Rule Based Approach

Day of week	Time of Day	Distance	Commute Time in Mins
Monday	10 AM - 11 AM	2 - 4 KM	40
Sunday	9 AM - 10 AM	4 - 6 KM	20
Tuesday	5 PM - 6 PM	2 - 4 KM	30
Friday	6 PM - 7 PM	4 - 6 KM	60

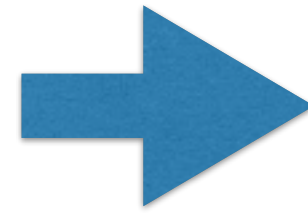


Rules like this are Static

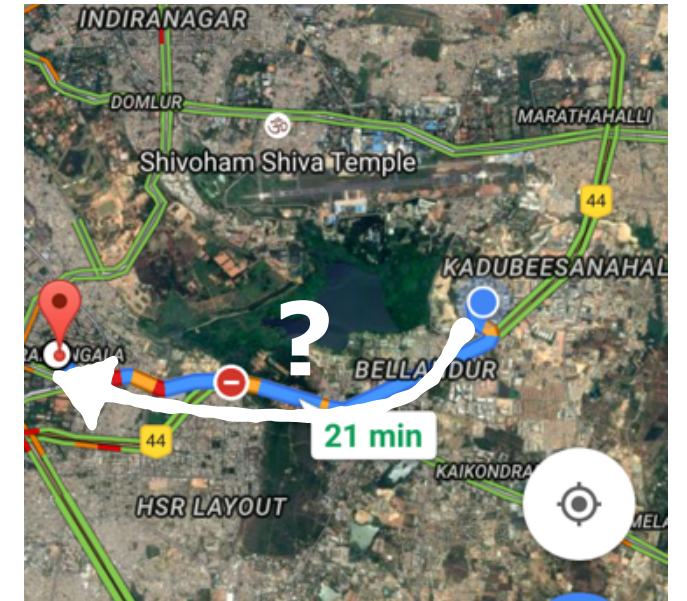
- They change really slowly over time
- Any changes are made by human analysts

Rule Based Approach

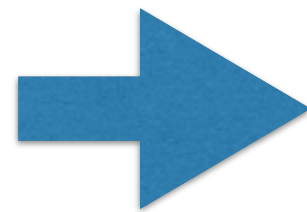
Day of week	Time of Day	Distance	Commute Time in Mins
Monday	10 AM - 11 AM	2 - 4 KM	40
Sunday	9 AM - 10 AM	4 - 6 KM	20
Tuesday	5 PM - 6 PM	2 - 4 KM	30
Friday	6 PM - 7 PM	4 - 6 KM	60



Static



**Traffic patterns
on the other hand**



Dynamic

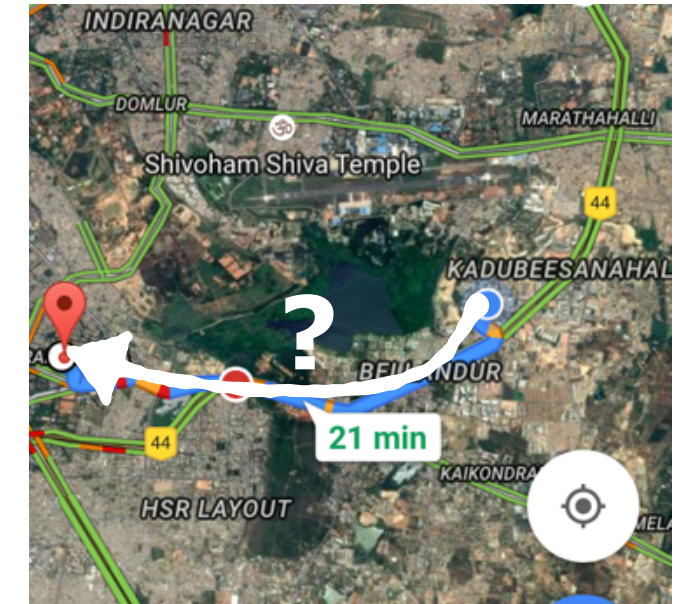
Two Approaches

Rule Based

**Machine
Learning Based**

Machine Learning Based Approach

- Collect a large amount of traffic data
- Use an algorithm to identify the relationship between the data and commute time
- Update this relationship continuously with new data

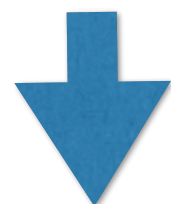


Rule Based Approach

Current Context (Source,
Dest, Time of day etc)



**Static
Rules**



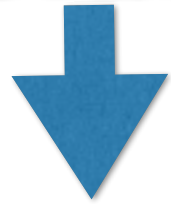
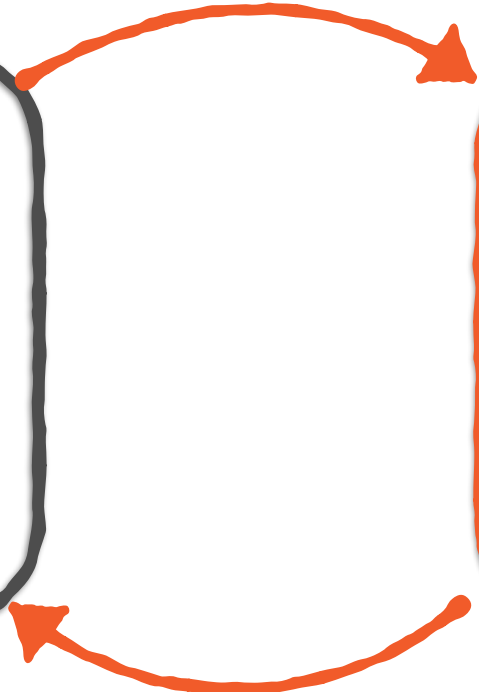
Commute Time

ML Based Approach

**Rules are updated
automatically
based on data**

ML Based Approach

Current Context (Source,
Dest, Time of day etc)



Commute Time

When to Use Machine Learning

- **Difficult for humans to express rules**
- **A large amount of historical data is available**
- **Patterns/Relationships are dynamic**

Typical ML Workflow

Pick your Problem

Identify which type of problem we need to solve

Represent Data

Represent data using numeric attributes

Apply an Algorithm

Use a standard algorithm to find a model

**Pick your
Problem**

**ML problems generally
fall under a broad set
of categories**

Classification

Regression

Recommendations

Clustering

**Pick your
Problem**

**Each type of problem has
its own basic workflow**

- How to set up the problem statement
- How to represent data

Typical ML Workflow

Pick your
Problem

Identify which type of
problem we need to
solve

Represent Data

**Represent data using
numeric attributes**

Apply an
Algorithm

Use a standard
algorithm to find a
model

Represent Data

Data might be in the form of Unstructured text, Images, Videos

Use meaningful numeric attributes to represent them

Typical ML Workflow

Pick your
Problem

Identify which type of
problem we need to
solve

Represent Data

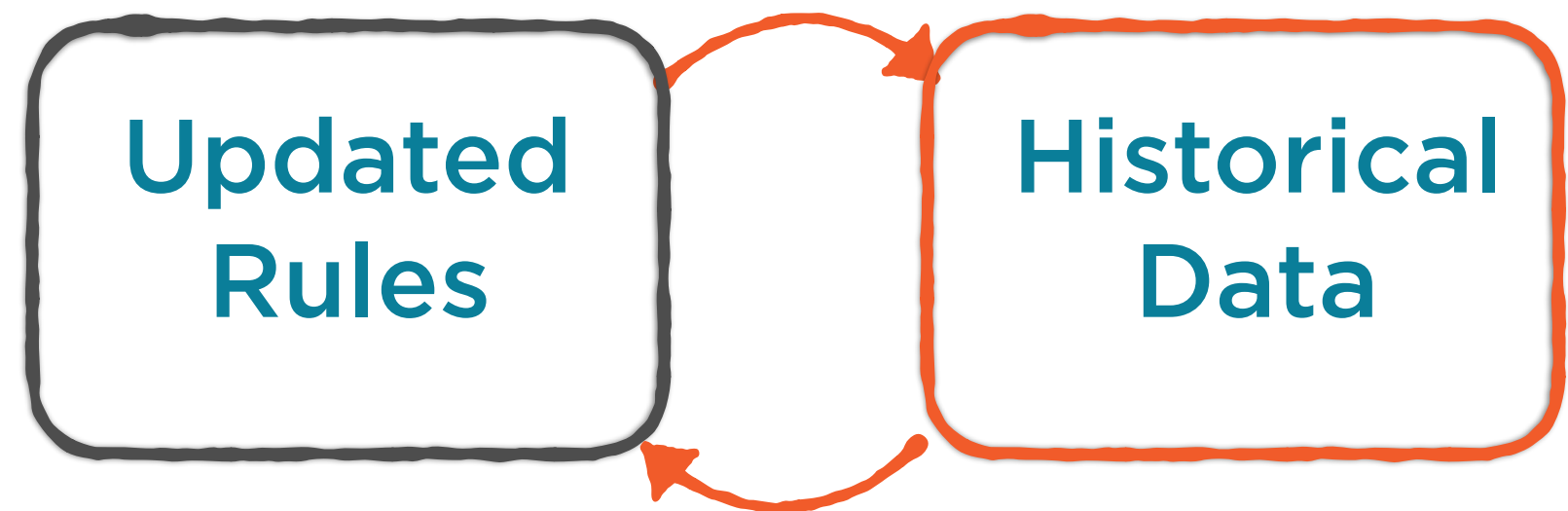
Represent data using
numeric attributes

Apply an
Algorithm

**Use a standard
algorithm to find a
model**

**Apply an
Algorithm**

**Use an algorithm to find
patterns from the historical
data**



**Apply an
Algorithm**

**Updated
Rules**

**Rules are meant to
quantify relationships
between variables**

**Apply an
Algorithm**

**Updated
Rules**

**The rules together
form something
called a Model**

Apply an
Algorithm

Model

A Model can be

- a mathematical equation
- a set of rules (if-then-else statements)

**Apply an
Algorithm**

**The choice of algorithm depends
mainly on the type of problem**

Classification

Naive Bayes

**Support Vector
Machines**

**Apply an
Algorithm**

**The choice of algorithm depends
mainly on the type of problem**

Clustering

K-Means

**Hierarchical
Clustering**

Typical ML Workflow



**This is usually
plug and play**

Typical ML Workflow



The diagram illustrates a typical machine learning workflow. It consists of three colored rectangular boxes arranged horizontally: a blue box on the left, an orange box in the middle, and a green box on the right. The first two boxes are enclosed within a dashed blue border. The text inside the boxes is white. Below the boxes, there is a line of text in a dark gray font.

Pick your
Problem

Represent Data

Apply an
Algorithm

You need to spend your
time and energy here

Typical ML Workflow



The diagram illustrates a three-step machine learning workflow. The first step, 'Pick your Problem', is highlighted with a blue background and a dotted blue border. The second step, 'Represent Data', has an orange background. The third step, 'Apply an Algorithm', has a green background. All steps are contained within a larger dotted blue border.

Pick your Problem

Represent Data

Apply an Algorithm

Picking the problem
involves making a number
of thoughtful choices

Types of ML Problems

Classification

Regression

Clustering

Recommendations

Classification

Spam Detection

Is this email **Spam** or **Ham**?

Sentiment Analysis

Is this tweet **positive** or **negative**?

Trading Strategy

Is this trading day going to be an **up-day** or **down-day**?

Classification

We are given a
problem instance

An e-mail

A Tweet

A trading day

Classification

We need to assign a category to the problem instance

Spam or **Ham**?
positive or **negative**?
up-day or **down-day**?

Classification

Algorithms which perform
classification are known as
Classifiers

Classification

A Classifier

uses a set of instances for which the correct category membership is known

Training Data

Ex: Tweets which are correctly classified as positive or negative

Types of ML Problems

Classification

Regression

Clustering

Recommendations

Regression

What will be **the price** of this stock on a given date?

How **long** will it take to **commute** from point A to point B?

What will be the **sales** of this product in a given week?

Regression

Compute a continuous value

Stock Price

Commute Time

Sales

Regression

You know the value
depends on certain inputs

Commute Time

depends on

Time of Day

Distance

Regression



Use
Regression to
identify this
function

Time of day,
Distance



Some
Function



Commute Time

Regression

Like Classification

Regression requires
Training Data

Ex: Historical datapoints

Date	Time of Day	Distance	Commute Time
------	-------------	----------	--------------

Types of ML Problems

Classification

Regression

Clustering

Recommendations

Clustering

Say you have a large group of users for a Social Network

Divide the users into **groups** based on some **common attributes**

Clustering

The key thing here is that..

..the groups to be divided
into are **unknown**
beforehand

Clustering

**The algorithm divides
users into groups**

**Later, we might realize that these groups
represent meaningful divisions**

**Likes, dislikes
Demographics**

Types of ML Problems

Classification

Regression

Clustering

Recommendations

Recommendations

What kind of **artists** will
this user **like**?

What are the **top 10**
book picks for this
user?

If a user buys this
phone, **what else** will
they buy?

Recommendations

Based on a user's past behavior

**Determine what else
they might like or need**

Collaborative Filtering

Typical ML Workflow



The choice made here will completely determine what happens in the next steps

Summary

Spot applications of Machine Learning in the real world

Differentiate between the different types of Machine Learning problems

Pick your problem: Classify, Regress, Recommend or Cluster