|  |  |  |
| --- | --- | --- |
| **Name** | **Amt** | **Fraud** |
| Alan | 1000 | No |
| Jerry | 2000 | Yes |
| Jamey | 2100 | Yes |
| Hill | 5000 | No |

Less data: Pattern that machine can understand is Names starting with J are Fraud

More data and more fields of data are better for machine learning

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Amt | Issued | Used | Age | Fraud |
|  |  |  |  |  |  |

So there is a pattern here

If amt > 5000

If age near 20

If card issued in one state used in other

Data

(Contains pattern)

Model

(Code/Logic to recognize pattern)

Machine learning Alogrithm

(Finds pattern)

Application uses Model to supply new data and see it matches known pattern

Like loading trx data for CC in and model can help understand if fraud or not

So what we need for it

1.Data

2.Compute

3.Machine L Algos

Who is data scientist

Some one who knows Stats/ML Algos/Domain

Rapid miner = Analytic vendor

Azure

Amazon = Cloud

R = open source for ML

Python

Machin L process

1.Iterative process

2.Challenging process

To solve a ML problem

Asking right question most imp part of process

See if the data is right to answer this question

Do you know how we validate the predictions from models

Raw data -> (Iterative: Preprocess this data) -> prepared data -> apply algo(iterate) -> model->deploy chosen model -> chosen model -> application use it

Need to recreate models as the world changes so needs to be iterative

Terminology

1. Training Data
   1. Prepared data used to create a model
   2. Creating a model is called training a model
2. Supervised Learning (Most commmon)
   1. Value you want to predict is in the training data
   2. Such data is called as labelled data
3. Unsupervised Learning
   1. Value you want to predict is not in the training data
   2. The data is un labelled

Preprocessing data (Iterative): Data curation and processing

1.Data from data sources is loaded to Data Preprocessing Moodules

2.Output of 1 is trainging data

Data has coloums

Coloums are called features

Value we trying to predict is also in training data

We call it Target value

Cateogaries of ML

1.Regression: Like how many units will we srll next month

2.Classification: Data Grouped into classes

Ex ques : is this credit card fraudulent (return prob)

3.Clustering: Find clusters in data (Unsupervised as no label data)

Ex ques: what are our customer segments

(we don’t know this but we can use data to predict it)

Styles of ML Algo

1.Decision Tree

2.neural nw

3.Bayesian

4.k-means

Trainign and Testing a model with supervised

1. Choose features to be most effective of the target value
   1. Contry issued
   2. Used
   3. Age
2. Input this data to learning algo (only 75% data)
3. Genarates a candidate model
4. Testing:
   1. We input tramining 25% data to predict if model is success
   2. Just compare the values which are already in traigning data
5. If not improving:
   1. Choosing different features
   2. Add more data
   3. Try changing different Algo

Using a model

* 1. List of movie recommendation

Logic and mathematics to create decision support systems

Philosophy to understand the nature of conscience, ethics, and human beings

Psychology to improve our understanding of how cognitive tasks may be performed (both by humans and machines)

Sociology to understand human-machine interactions

Language theory to understand how semantics can be captured and evolves

Research is been done to mimic human intelligence with machines and will continue to be a hot trend topic in future

There can be a day when the AI will outperform usin some of the tasks that we require a human for doing

However AI is not a new thing and we have been using it since past

Remember decision support systems or rule based systems used in early days of software era and are still being used now

Many of us are not aware that we are using AI technology when searching the Internet, using our credit card, or chatting with a call center guys. AI technology is all around us.

AI has 2 types

1. Deterministic (Logic/Computation)
2. Non Deterministic (Supervised and Non Supervised)

Deterministic AI

1. This is a rule based intelligence
2. Here we have a set of rules defined and if the conditions fit the rule then we can use the decision that the rule outcomes
3. Actually speaking its not intelligent and do not produce any outcomes for unknown data
4. It cannot deal with uncertainty
5. Result is always a certain solution - same outcome for same input

Non Deterministic AI

1. This uses a rule based approach to minimize or zero down to a conclusions but more potentially uses statistics and bigdata to reach to a result
2. Here the outcome is unknown like
   1. In case of let say a credit card transaction we don’t know if the user is a fraud person in advance
   2. Or if it’s a congestion management system we don’t know how people will react in case of a traffic congestion
3. In such cases machine may use a set of algorithms to derive the pattern and may help in predicting the result
4. Results are uncertain - multiple outcomes for same case
5. Makes use of machine learning to get to a prediction

In short AI is to make a machine intelligent enough so it can assist a human in making decisions may be simple/complex

Now to make a machine intelligent it has to learn

This learning is done using a something called as Machine Learning

Machine Learning is to teach a machine to do

Non Deterministic AI uses ML that has two types

1. Supervised ML

Model

(Code/Logic to recognize pattern)

ML Algorithm

(Finds Pattern)

Data

(Contains Pattern)

Application

How it works

1. Data from the system is preprocessed and then fed to ML Algorithm
2. ML Algorithm finds a pattern based on the features in data and produces a Model
3. Model is something like a logic/code that can take in new data and try to recognize the pattern in it to return a prediction

generate results for a task based on patterns that are found using advanced statistics and (big) data. The model is created using a generic method, and the resulting input-output relationship is implicit. Typically, these methods generate multiple outcomes for the same state or input.

The result could be incorrect. Outcomes are, therefore, rated based on measures such as accuracy, generality, and precision. Some results are rated better than other results and, depending on the problem domain, we may be satisfied with anything better than random accuracy or an accuracy level close to 100%

Superwised - algo are taught by set of rules

Un - Not taught

UN - Neural Networks = Algos and Deeplearning = combn of NN

XAI = Explanable Ai

WIRE MOCK