python

Class: Machine Learning

Topic

Introduction to Machine Learning

What We Have Learnt

Getting started with exploratory data analysis

Machine Learning

What We Will Learn

Introduction

Examples

Tasks Performed

Algorithms

Advance Concepts

Implementation Using Credit Default Dataset

Getting Started

What is Machine Learning?

"Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed"

- **Arthur Samuels, 1959**

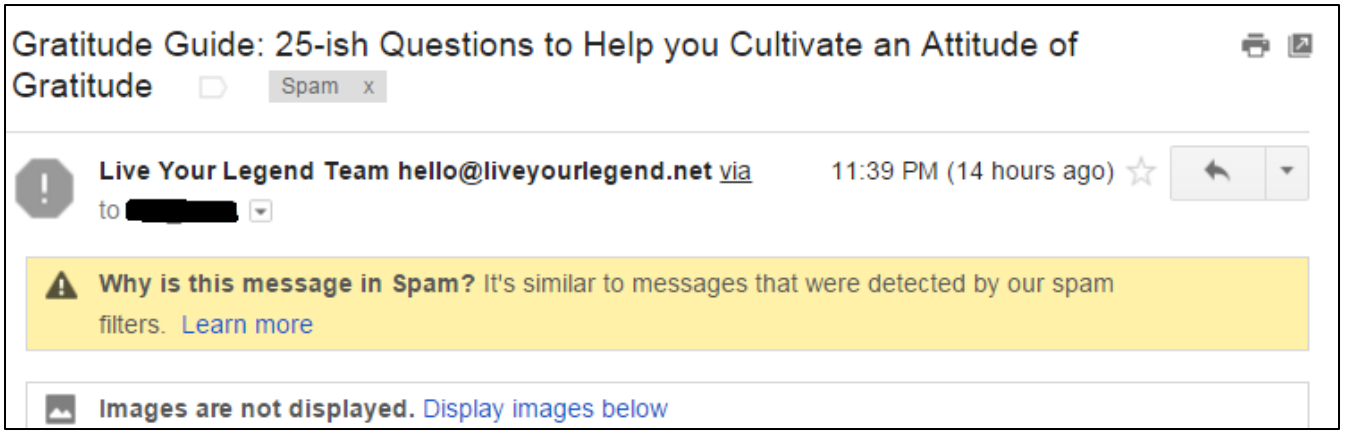
"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

- **Tom Mitchell, Carnegie Mellon University**

Examples

Our Daily Lives

Gmail Spam Filtering



Any popular e-mail service provider today can mark e-mails as spam with a reasonable degree of accuracy

Examples

Our Daily Lives

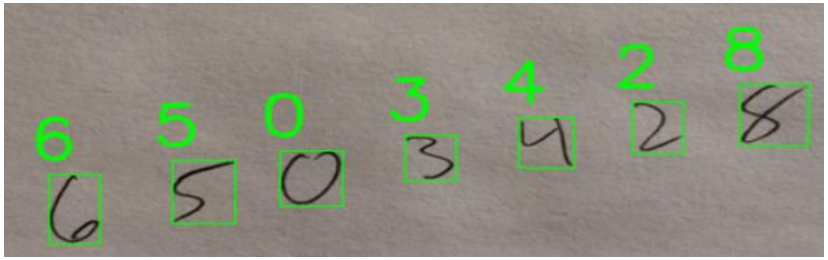
Product Recommendation by Ecommerce Websites

Application of machine learning algorithms

Examples

Our Daily Lives

Handwritten Character Recognition



Machine learning algorithms can classify handwritten digits with close to 100% accuracy

Examples

Our Daily Lives

ABS on Cars

Auto pilot on Tesla cars

Self-driving cars from Google

Evolution of Machine Learning

Ideas and Methodologies

Probability and Statistics

Computer Science

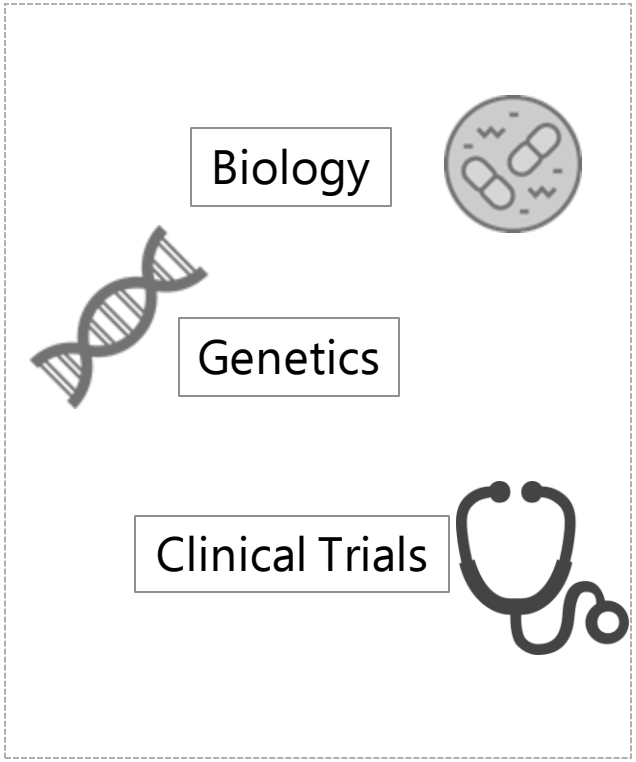
Computational Biology

Backbone of modern Machine Learning theory

Evolution of Machine Learning

Ideas and Methodologies

Fields from Social Sciences

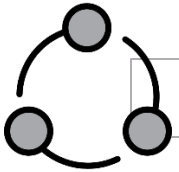


Evolution of Machine Learning

Ideas and Methodologies

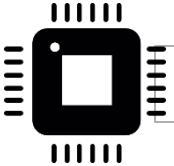
Advent of Modern Computing Power

Credit Card



Cluster from any of the available web services

Basic to advance specifications



Ensures enterprise scale hardware availability for consumers

Evolution of Machine Learning

Ideas and Methodologies

Good results on problems considered historically difficult to solve have given Machine Learning much popularity in Press

Getting Started with Machine Learning

Starting point should be a question with a specific business context in mind

Getting Started with Machine Learning

Why is Image Recognition Important?

Helps solve a specific business problem

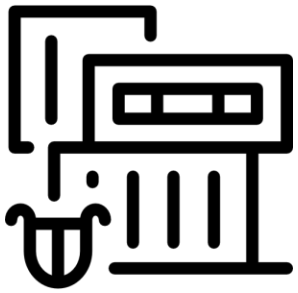
Getting Started with Machine Learning

The Process

Why do you think predicting whether a customer will default credit card payment next month is useful from a business perspective

Getting Started with Machine Learning

The Process

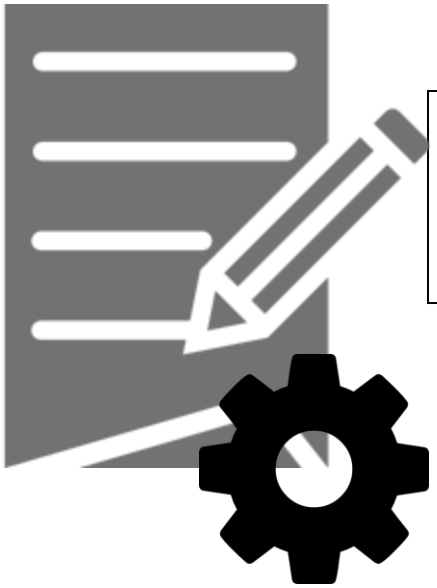


Data relevant to the question needs to be gathered

Historical data helps the machine learn from it

Getting Started with Machine Learning

The Process



Once the machine learns from historical data, it needs some additional data to determine how well it has learnt

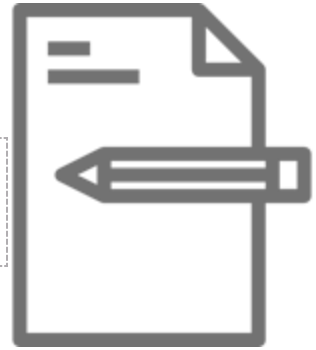
Getting Started with Machine Learning

The Process

Example

Probability/This course

There is classroom material that helps the students learn the concepts along with some real life examples



Once classroom lectures are done, students get evaluated based on an examination at the end of the course

Questions may not come from the materials that were studied

Getting Started with Machine Learning

The Process

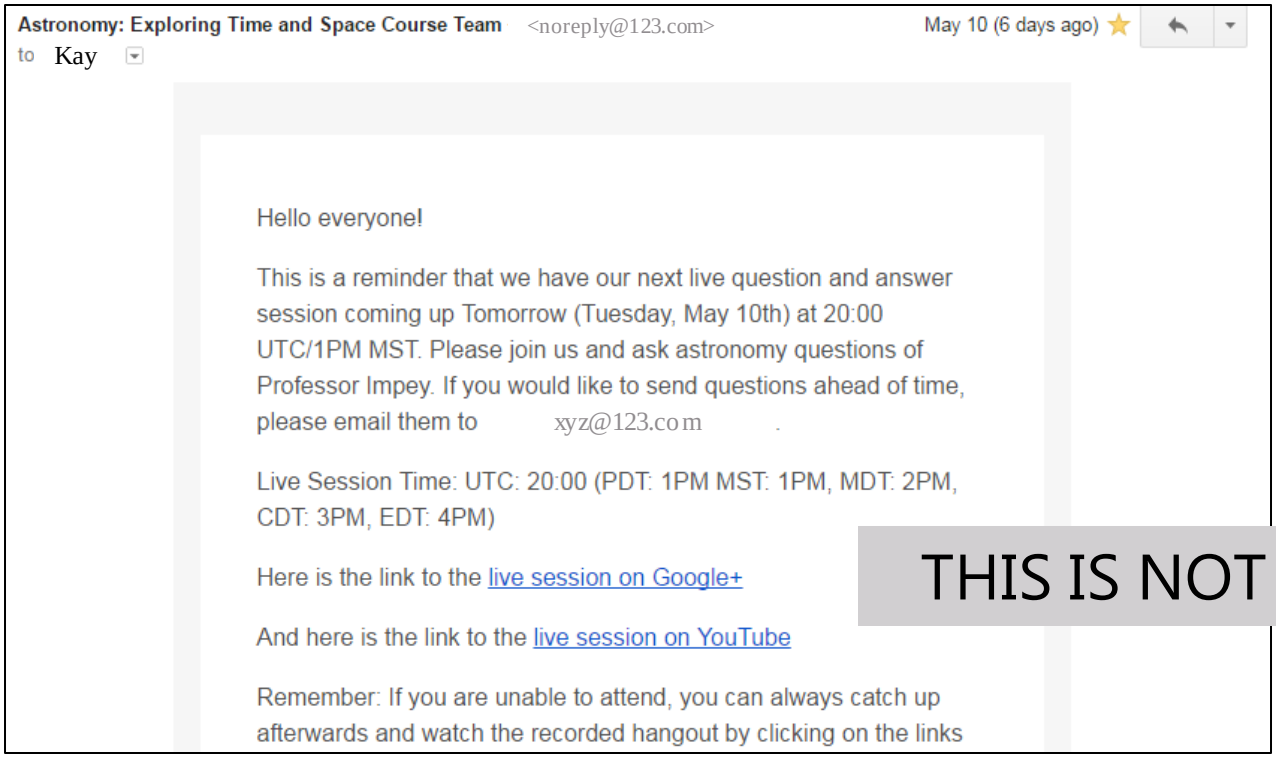
Data

Raw data comes in different formats

Getting Started with Machine Learning

The Process

Data

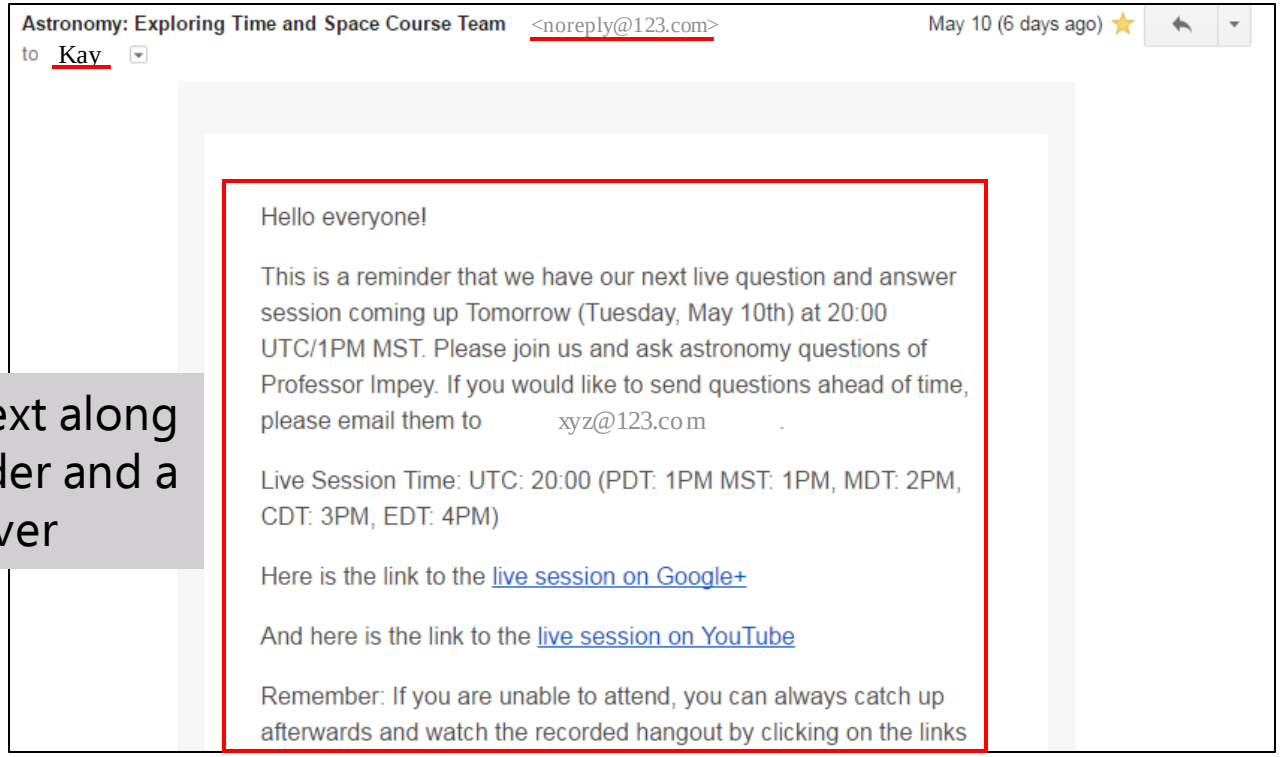


THIS IS NOT SPAM!

Getting Started with Machine Learning

The Process

Data

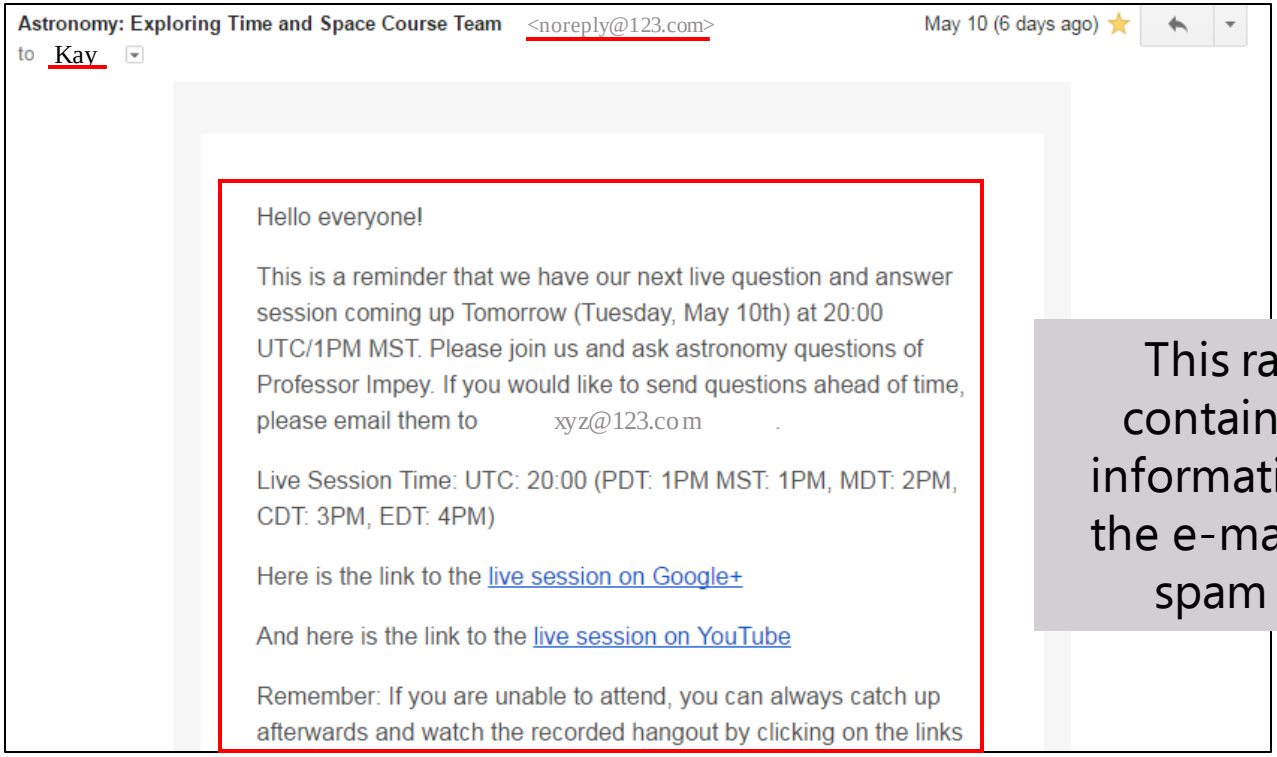


Bunch of text along with a sender and a receiver

Getting Started with Machine Learning

The Process

Data

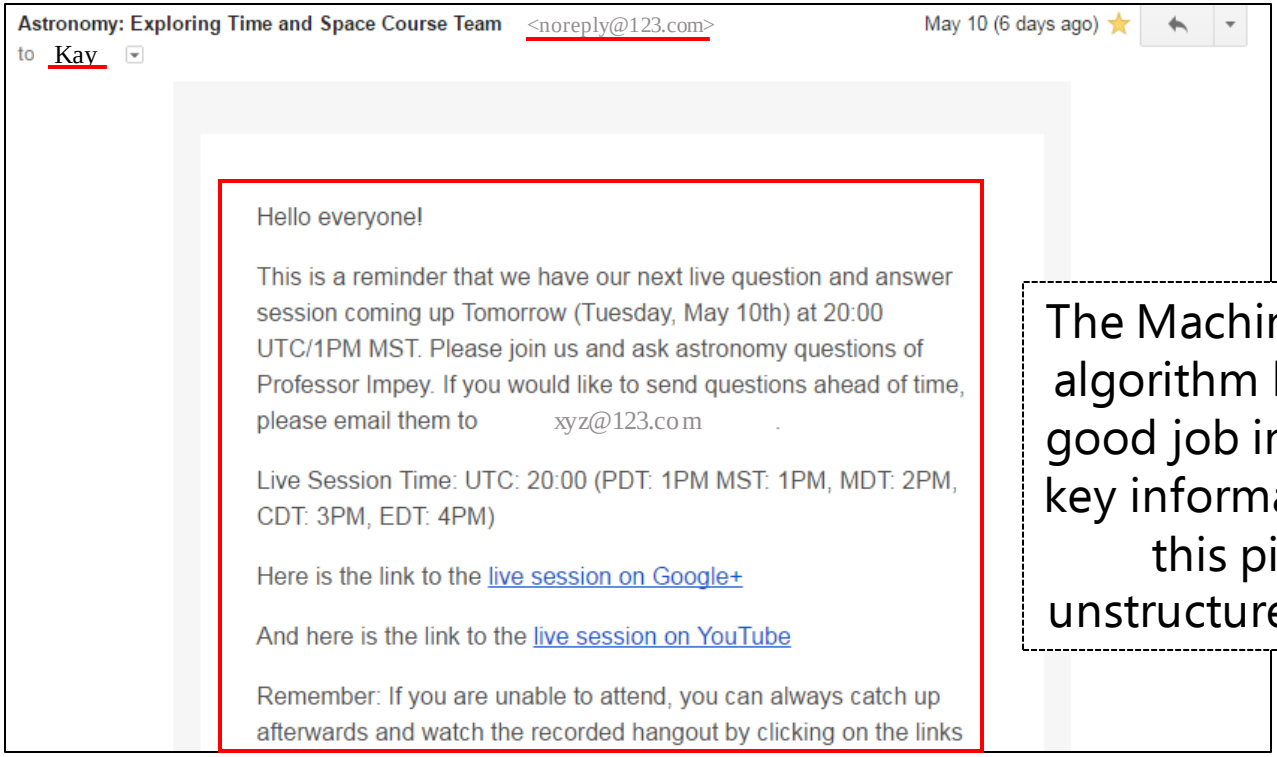


This raw data contains useful information about the e-mail being a spam or not

Getting Started with Machine Learning

The Process

Data

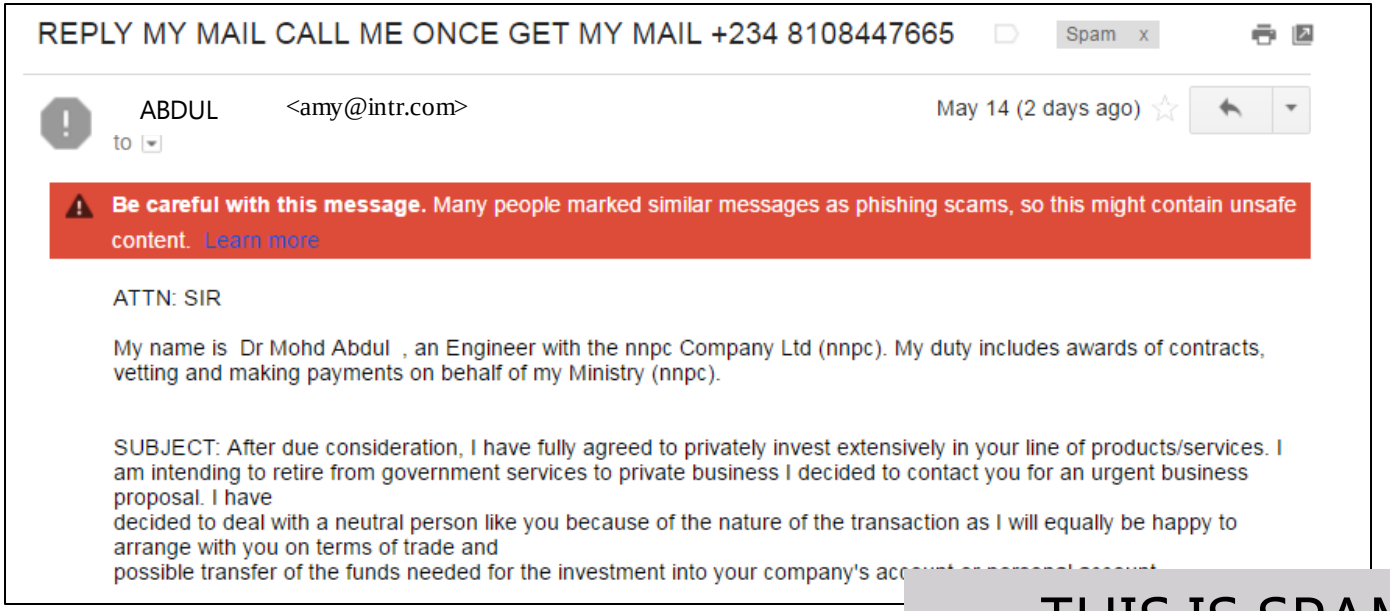


The Machine Learning algorithm has to do a good job in extracting key information out of this piece of unstructured raw data

Getting Started with Machine Learning

The Process

Data


REPLY MY MAIL CALL ME ONCE GET MY MAIL +234 8108447665 Spam x

ABDUL <amy@intr.com> May 14 (2 days ago)

to

Be careful with this message. Many people marked similar messages as phishing scams, so this might contain unsafe content. Learn more

ATTN: SIR

My name is Dr Mohd Abdul , an Engineer with the nnpc Company Ltd (nnpc). My duty includes awards of contracts, vetting and making payments on behalf of my Ministry (nnpc).

SUBJECT: After due consideration, I have fully agreed to privately invest extensively in your line of products/services. I am intending to retire from government services to private business I decided to contact you for an urgent business proposal. I have
decided to deal with a neutral person like you because of the nature of the transaction as I will equally be happy to arrange with you on terms of trade and
possible transfer of the funds needed for the investment into your company's ac

THIS IS SPAM!

Getting Started with Machine Learning

The Process

Data

The sender, the e-mail text contain some useful information about the legitimacy of the e-mail

Getting Started with Machine Learning

The Process

How do we get a machine to extract information out of raw data that will help it learn a specific problem?

Feature Engineering

Conversion of raw data to useful machine readable informative features is called **feature engineering**

Feature Engineering

Going Back

During the exploratory analysis, the average bill amount for each customer over the last 6 months was taken – a form of feature engineering

Feature Engineering

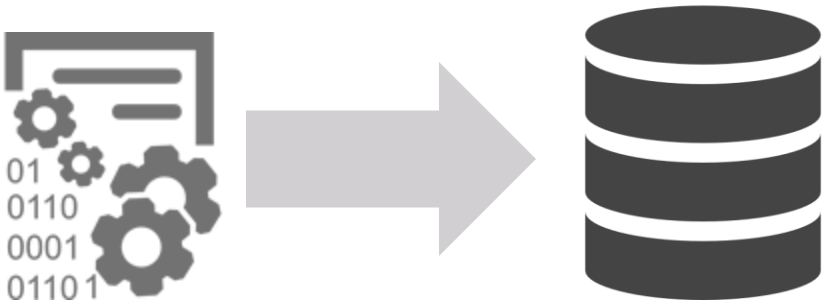
The number of features is only limited by creativity and imagination

How feature engineering is performed on a data set can have a profound impact on final results

Feature Engineering

Starting Point

Going from seemingly unstructured data to something that is more tidy, clean and structured



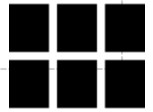
Feature Engineering

Starting Point

Print shape 

 Read data in Pandas

Credit Data 

29000 rows and 25 columns 

Row – customer
Column – information about the customer

Rectangular format
Tidy Data

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

3 e-mails

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

Whether or not the sender is in the receiver's contact list

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

x1 is a Boolean value – it can take one of 2 possible values – **True** or **False**

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

Counts the number of special characters in the e-mail, denoted by **x2**

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

This is also a Boolean which marks whether the e-mail is spam

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

This is the variable that needs to be predicted using Machine Learning algorithm

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

This variable is called a **response**

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

Also known as **Outcome** or **Dependent** variable

Feature Engineering

Example

Email No.	Sender in Contact List (x1)	Number of special Characters (x2)	Spam
1	YES	4	NO
2	NO	23	YES
3	YES	45	YES

Alternative terms for features are **independent variables**, **explanatory variables**

Feature Engineering

Exercise

Think of some more interesting features that you can construct from the raw data that might be useful in classifying an e-mail as spam or not

Recap

Introduction to Machine Learning

What is Machine Learning?

Examples

Evolution

Getting Started

Feature Engineering

Next

Types of Tasks, Machine Learning Algorithms and Linear Regression