



NELLY OKTAVIA ADIWIJAYA, S.SI, MT, CEC  
FASILKOM UNIVERSITAS JEMBER

# DATA SCIENCE

Pendahuluan Data Science, Big Data dan Open Data

# Tujuan Pembelajaran

## **Pada kuliah ini anda akan:**

- Mengetahui orang-orang yang bekerja di bidang data science
- Memahami definisi data science
- Mempelajari tentang data science dalam konteks bisnis
- Menemukan beberapa kasus yang dapat menggunakan data science



# WHAT IS DATA SCIENCE

adalah proses menggunakan data untuk memahami berbagai hal, untuk memahami dunia.

Ketika Anda memiliki model, atau hipotesis dari suatu masalah, dan Anda mencoba untuk memvalidasi hipotesis atau model tersebut dengan data Anda.

**DATA SCIENCE IS A PROCESS, NOT AN EVENT**

# What is Data Science

**Mengungkap insight dan tren yang tersembunyi di balik data.**

Menerjemahkan data menjadi cerita, sehingga storytelling tersebut dapat membangun insight. Dan dengan insight tersebut, Anda dapat membuat pilihan strategis untuk sebuah perusahaan atau institusi.

Mengekstrak data dari berbagai bentuk, baik itu bentuk tidak terstruktur atau terstruktur.

# What is Data Science

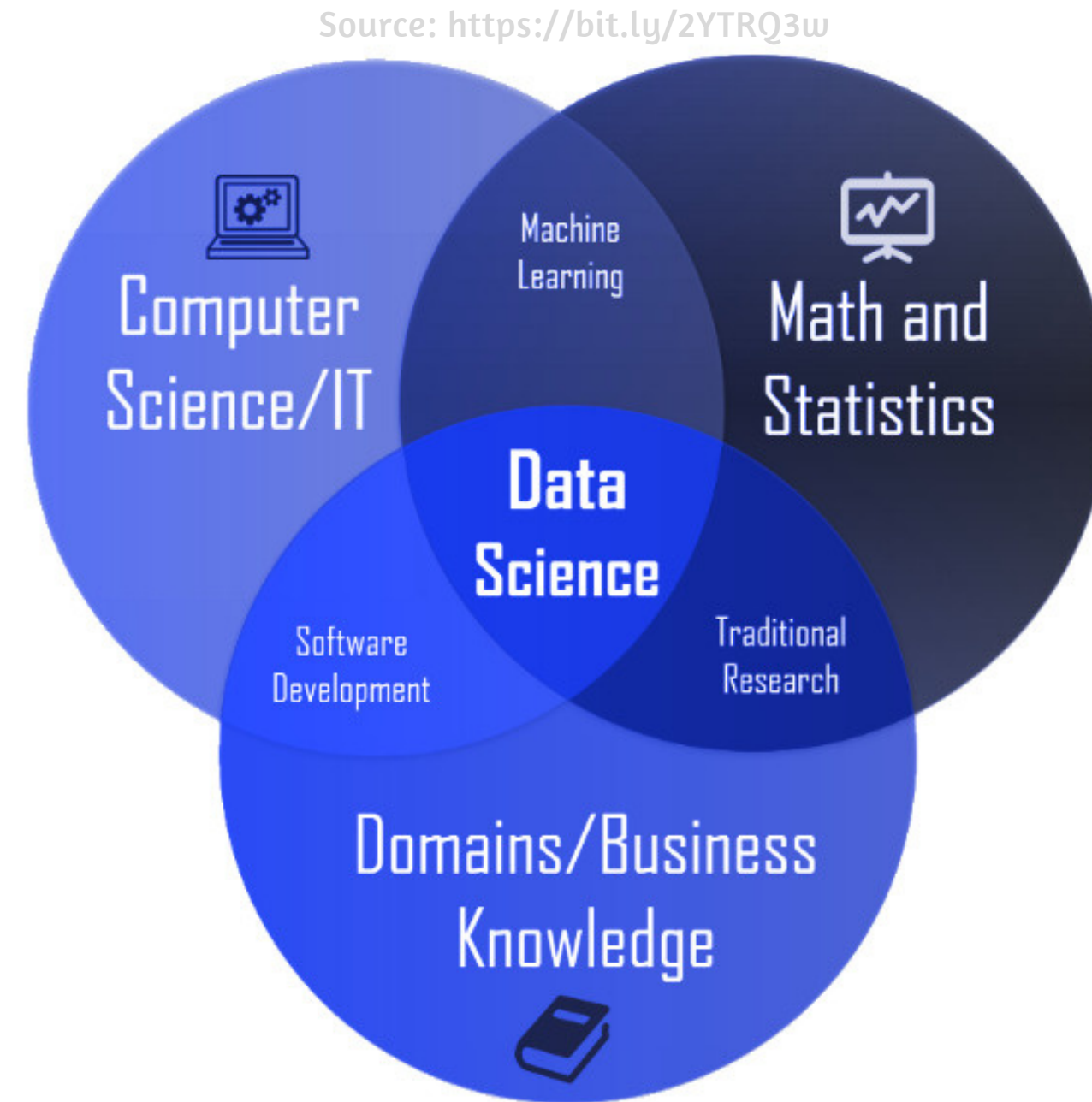
Data Science melibatkan data dan beberapa bidang ilmu. Definisi, atau namanya, muncul di tahun 80-an dan 90-an ketika beberapa profesor melihat ke dalam kurikulum statistik, mereka berpikir akan lebih baik menyebutnya Data Science.

# What is Data Science



bidang multi-disiplin yang menggunakan metode, proses, algoritma, dan sistem ilmiah untuk mengekstrak pengetahuan dan wawasan dari data terstruktur dan tidak terstruktur.

# What is Data Science



sebuah "konsep untuk menyatukan statistik, analisis data, machine learning, dan metode terkaitnya" untuk "memahami dan menganalisis fenomena aktual" dengan data. menggunakan teknik dan teori yang diambil dari berbagai bidang dalam konteks matematika, statistik, ilmu komputer, dan ilmu informasi



# Data Science Today

Data Science relevan saat ini karena banyaknya data yang tersedia

Dulu khawatir kekurangan data, saat ini data sangat melimpah

Di masa lalu, belum ada algoritmanya, sekarang banyak algoritma

Dulu software harganya mahal, sekarang open source dan gratis

Jika Anda memiliki data, memiliki rasa ingin tahu, dan Anda bekerja dengan data, serta memanipulasinya, Anda menjelajahnya, latihan menganalisis data, **mencoba mendapatkan jawaban darinya, adalah data Science**



# **WHO CAN BE A DATA SCIENTIST?**

- **Data Science belum ada hingga 2009, 2011.**
- **Sebelumnya hanya ada statistik**
- **Semua orang bisa menjadi IData Scientist**
- **Statistik, Matematika, Ilmu Komputer, Bisnis, Latar belakang Keuangan**
- **Orang-orang dari berbagai latar belakang bekerja sebagai Data Science sekarang**
- **Yang perlu Anda miliki hanyalah CURIOSITY**

# WHO CAN BE A DATA SCIENTIST?

Data scientist harus **ingin tahu, sangat argumentatif, menghakimi**.

**Rasa ingin tahu** mutlak diperlukan. Jika Anda tidak penasaran, Anda tidak akan tahu apa yang harus dilakukan dengan data tersebut.

**Menghakimi** karena jika Anda tidak memiliki prasangka tentang berbagai hal, Anda tidak akan tahu harus mulai dari mana.

**Argumentatif** karena jika Anda bisa berdebat maka Anda bisa mengajukan kasus, setidaknya Anda bisa mulai dari suatu tempat. Dan kemudian Anda belajar dari data dan kemudian Anda memodifikasi asumsi dan hipotesis Anda, dan data Anda akan membantu Anda belajar. Dan Anda mungkin mulai dari titik yang salah, Anda mungkin mengatakan bahwa saya pikir saya percaya ini tapi sekarang dengan data saya tahu ini, jadi ini memungkinkan Anda proses pembelajaran.

Hal terakhir yang dibutuhkan ilmuwan data dan itu adalah kemampuan untuk menceritakan sebuah Story.

Bahwa setelah Anda memiliki analisa, setelah Anda memiliki tabulasi, seharusnya sekarang Anda dapat menceritakan Story yang hebat.

Karena jika Anda **tidak menceritakan Story hebat darinya, temuan Anda akan tetap tersembunyi**, itu akan tetap terkubur, tidak ada yang akan tahu, dan anda akan cepat terkenal dengan cukup mengandalkan kemampuan Anda untuk menceritakan Story2 hebat.

# Roles Required in Data Science Project

- Prove / disprove hypotheses.
- Information and Data gathering.
- Data wrangling.
- Algorithm and ML models.
- Communication.

**Data Scientist**



- Build Data Driven Platforms.
- Operationalize Algorithms and Machine Learning models.
- Data Integration.

**Data Engineer**



- Storytelling.
- Build Dashboards and other Data visualizations.
- Provide insight through visual means.

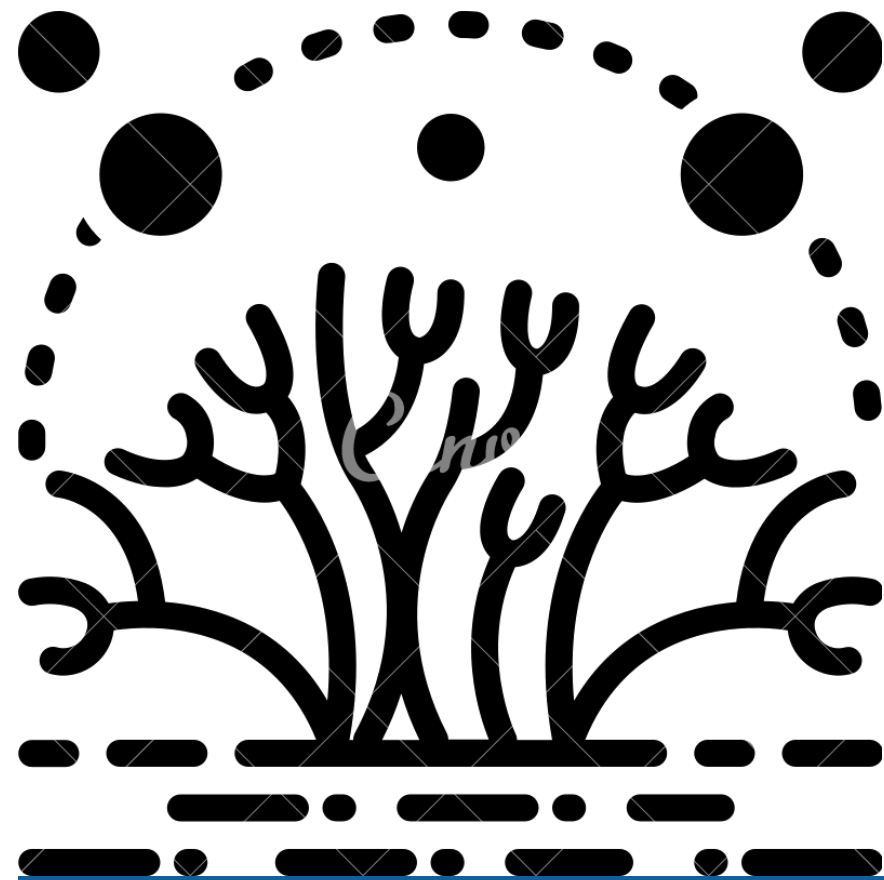
**Visualization Expert**



- Project Management.
- Manage stakeholder expectations.
- Maintain a Vision.
- Facilitate.

**Process Owner**





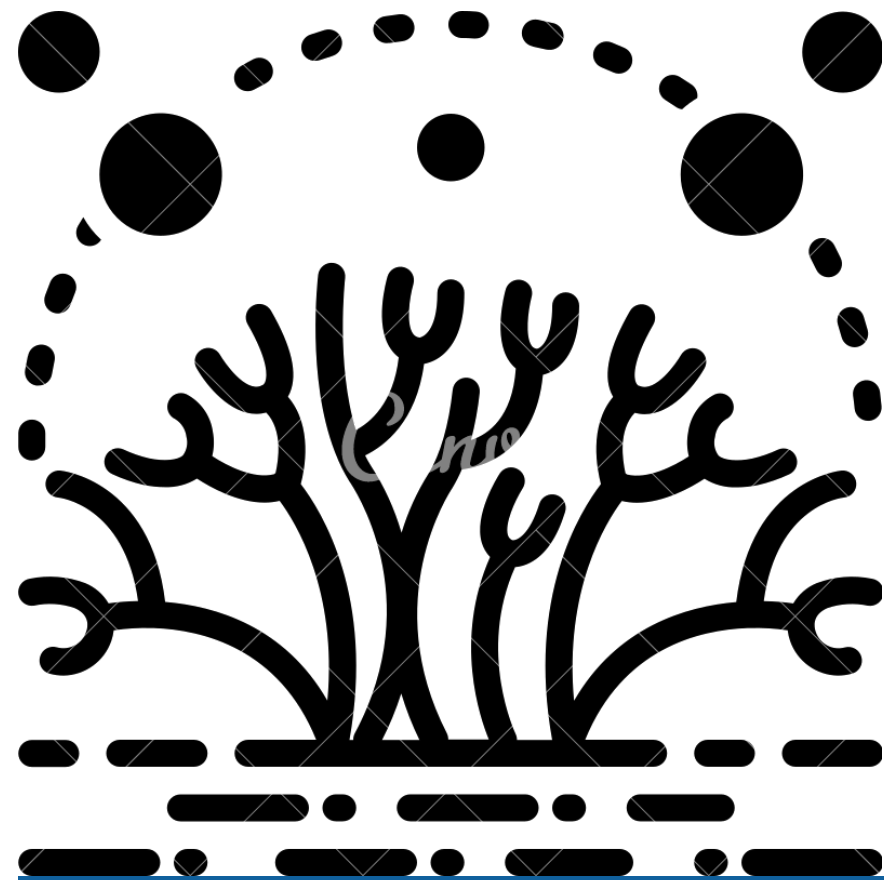
## What do Data Science people do? (Example)

Contoh:

di universitas jurusan teknik kimia terdapat masalah yaitu ingin memprediksi ganggang mekar.

Mekarnya alga ini dapat menyebabkan peningkatan toksisitas air dan dapat menyebabkan masalah pada perusahaan pengolahan air.

Hal itu tidak dapat diprediksi dengan latar belakang teknik kimia.




## What do Data Science people do? (Example)

Solusi :

Dengan menggunakan **jaringan saraf tiruan** dapat memprediksi kapan perkembangan alga ini akan terjadi.

Sehingga perusahaan pengolahan air dapat menangani masalah ini dengan lebih baik.



## What do Data Science people do? (Example)

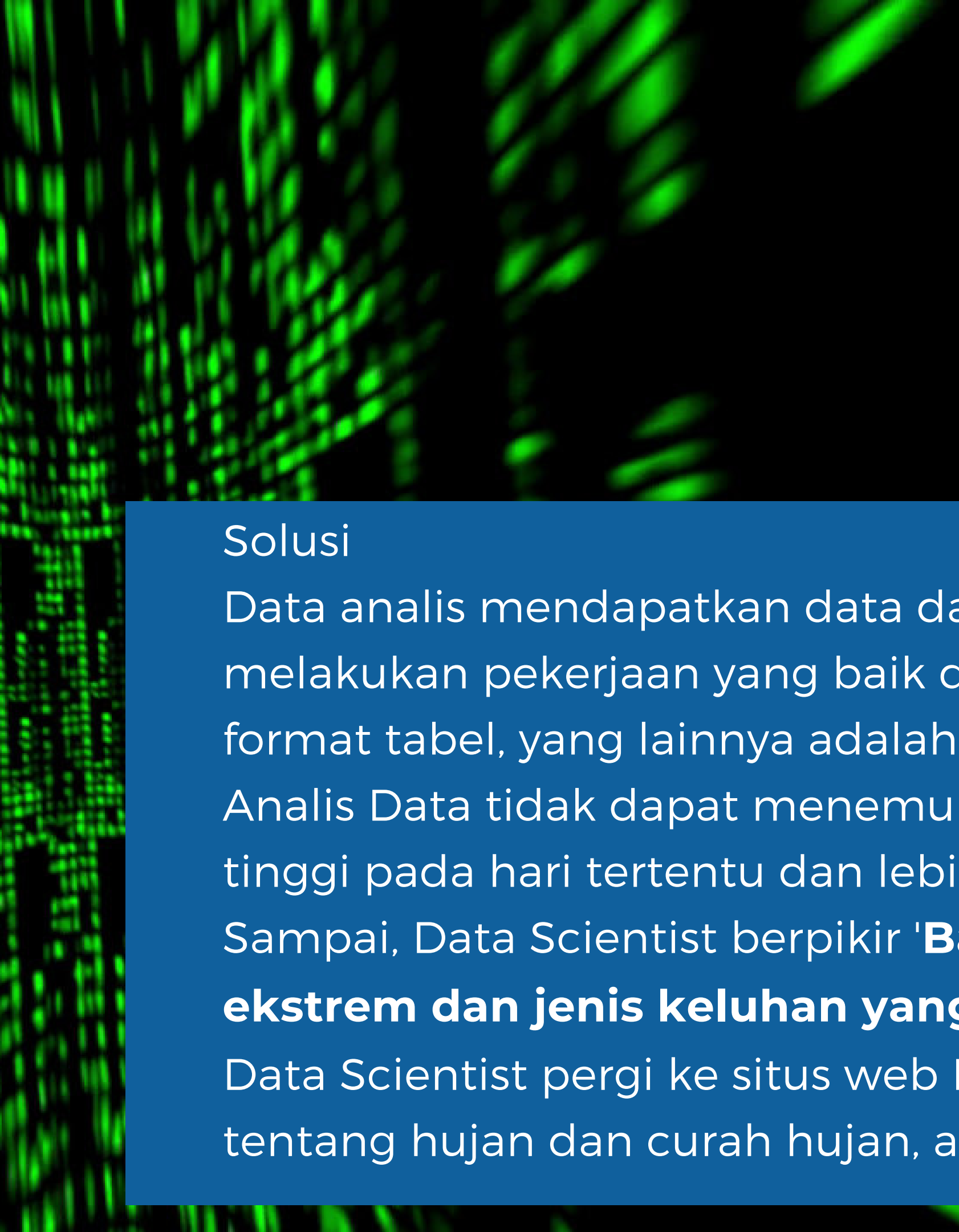
Masalah

Di Toronto angkutan umum dioperasikan oleh Toronto Transit Commission (TTC). Salah satu otoritas transit terbesar di wilayah Amerika Utara.

Dan suatu hari mereka menghubungi Data Scientist dan berkata, 'kami memiliki masalah, **kami memiliki data keluhan dan ingin menganalisisnya** dan kami membutuhkan bantuan Anda'. Mereka berkata, 'beberapa'.

Berapa banyak keluhan? 'Mungkin setengah juta'.





## What do Data Science people do? (Example)

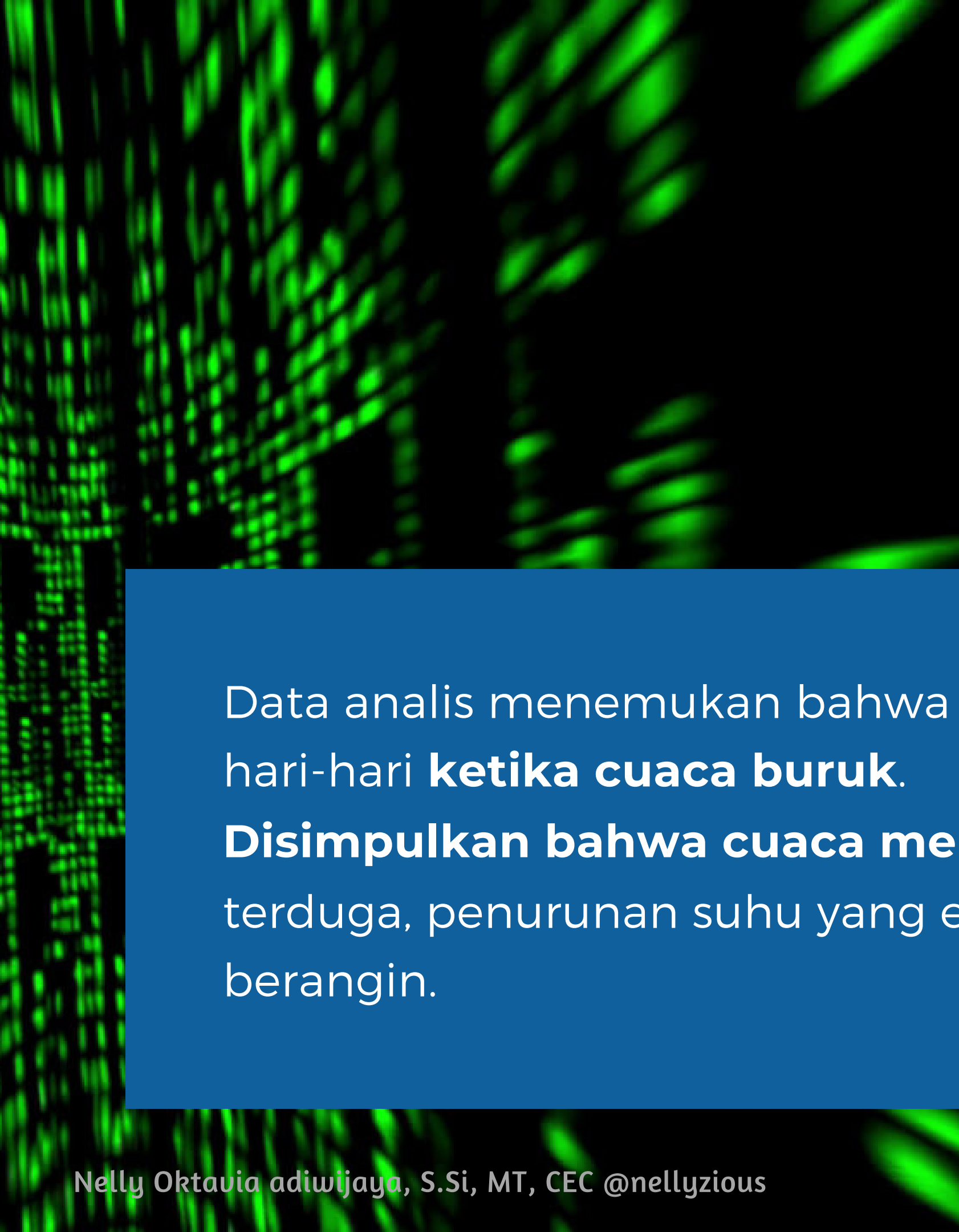
### Solusi

Data analis mendapatkan data dan mulai menganalisisnya. Pada dasarnya TTC telah melakukan pekerjaan yang baik dalam menyimpan data, beberapa data dalam format tabel, yang lainnya adalah data tidak terstruktur.

Analisis Data tidak dapat menemukan motif yg menjadikan keluhan yang satu lebih tinggi pada hari tertentu dan lebih rendah pada hari lainnya.

Sampai, Data Scientist berpikir '**Bagaimana jika ada hubungan antara cuaca ekstrem dan jenis keluhan yang diterima TTC?**'

Data Scientist pergi ke situs web Environment Canada dan mendapatkan data tentang hujan dan curah hujan, angin dan sejenisnya.



## What do Data Science people do? (Example)

Data analis menemukan bahwa sepuluh hari yang paling banyak keluhan, adalah hari-hari **ketika cuaca buruk**.

**Disimpulkan bahwa cuaca menyebabkan keluhan.** yaitu hujan yang tidak terduga, penurunan suhu yang ekstrem, terlalu banyak salju, hari yang sangat berangin.

# Data Science in Business

Dalam bisnis, **jika mereka tidak dapat mengukur sesuatu maka mereka tidak dapat memperbaikinya** (improve)

Dan jika tidak dapat mengukur biaya mereka, mereka tidak dapat mengurangnya.

Jika mereka tidak dapat mengukur keuntungan mereka, mereka tidak dapat meningkatkannya.

Jadi hal pertama yang harus dilakukan perusahaan adalah **mulai merekam informasi, mulai menangkap data.**

Data tentang biaya kemudian dibedakan menjadi biaya tenaga kerja dan biaya material.

Masih di Biaya, berapa biaya untuk menjual satu produk dan berapa total biaya.

# Data Science in Business

Dan kemudian Anda melihat pendapatan. Dari mana pendapatan Anda berasal?

---

Apakah 80% pendapatan Anda berasal dari 20% pelanggan Anda? Atau justru sebaliknya?

---

**Jadi hal paling pertama, mulai menangkap data. Setelah Anda memiliki data, Anda dapat menerapkan algoritme dan menganalisanya.**

---

Jadi hal pertama yang harus dilakukan adalah mengambil data. Jika Anda tidak menangkapnya, mulailah menangkapnya. Jika Anda merekamnya, Arsipkan.

# RECRUITING FOR DATA SCIENCE IN BUSINESS

Orang yang mereka pekerjakan, mereka harus **senang bermain dengan data**, dan kemudian mereka tahu cara bermain dengan visualisasi data.

**Mereka memiliki pemikiran analitis.**

**Tapi yang benar-benar penting adalah yang memiliki passion tentang jenis bisnis Anda.**

Seseorang bisa menjadi ilmuwan data yang hebat di lingkungan ritel, tetapi mereka mungkin tidak begitu tertarik untuk bekerja di perusahaan yang terkait dengan TI, atau bekerja dengan log web berukuran gigabyte.

Tetapi jika seseorang tertarik dengan log web tersebut, atau seseorang tertarik dengan hubungan data, maka mereka akan dapat berkontribusi lebih besar pada produktivitas Anda.

# Use Case of Data Science

**Bagaimana perusahaan dapat menggunakan semua informasi** yang mereka kumpulkan dari pelanggan mereka untuk benar-benar mengembangkan produk baru yang menanggapi kebutuhan pelanggan?

Google Search adalah aplikasi Data Science. **Google Search, kita bisa mencari apapun kapanpun.** Apapun Google sekarang, itu semua karena Data Science.

Aplikasi yang sempat booming : Pokemon Go.  
Mereka menggunakan data aplikasi Ingress. Aplikasi terakhir dari perusahaan yang sama dan mereka yg memilih lokasi untuk Pokemon dan gym berdasarkan data dari aplikasi terakhir. Jadi mereka belajar dengan kesalahan mereka.



# Data Science with Open Data

## What is Open Data

**Open Data didefinisikan sebagai data terstruktur yang dapat dibaca mesin, dibagikan secara bebas, digunakan, dan dibangun tanpa batasan. Knowledge is Open jika ada orang yang bebas mengakses, menggunakan, memodifikasi, dan membagikannya.**



# DATA SCIENCE WITH OPEN DATA

**Definisi Open memberikan definisi yang lebih rinci tentang Open Data. Ringkasan poin paling penting:**

**Ketersediaan dan Akses:** data harus tersedia secara keseluruhan dan tidak lebih dari biaya reproduksi yang wajar, sebaiknya dapat diunduh melalui internet. Data juga harus tersedia dalam bentuk yang nyaman dan dapat dimodifikasi.

**Penggunaan kembali dan Distribusi Ulang:** data harus disediakan di bawah persyaratan yang mengizinkan penggunaan kembali dan pendistribusian ulang termasuk pencampuran dengan kumpulan data lainnya.

**Partisipasi Universal:** setiap orang harus dapat menggunakan, menggunakan kembali, dan mendistribusikan kembali. Tidak boleh ada diskriminasi terhadap bidang usaha atau terhadap orang atau kelompok. Misalnya, pembatasan 'non-komersial' yang akan mencegah penggunaan 'komersial', atau pembatasan penggunaan untuk tujuan tertentu (misalnya hanya dalam pendidikan), tidak diperbolehkan.

# KEUNTUNGAN OPEN DATA

**Dukungan untuk inovasi** - Akses ke sumber daya pengetahuan dalam bentuk data mendukung inovasi di sektor swasta dengan mengurangi duplikasi dan mempromosikan penggunaan kembali sumber daya yang ada. Ketersediaan data dalam bentuk yang dapat dibaca mesin memungkinkan penggabungan kreatif yang dapat digunakan untuk menganalisis pasar, memprediksi tren dan persyaratan, dan mengarahkan bisnis dalam keputusan investasi strategis mereka.

**Memajukan akuntabilitas pemerintah dan reformasi demokrasi** – peningkatan akses ke data dan informasi pemerintah memberikan wawasan yang lebih luas kepada publik tentang kegiatan pemerintah, pemberian layanan, dan penggunaan uang pajak.

# KEUNTUNGAN OPEN DATA

## **Memanfaatkan informasi sektor publik untuk mengembangkan produk konsumen dan komersial**

Akses terbuka dan tidak terbatas ke data ilmiah untuk tujuan kepentingan publik, khususnya informasi statistik, ilmiah, geografis, dan lingkungan, memaksimalkan penggunaan dan nilainya, dan penggunaan kembali data yang ada dalam aplikasi komersial meningkatkan waktu ke pasar untuk bisnis.

## **Penggunaan yang lebih baik dari investasi yang ada di broadband dan infrastruktur informasi komunitas**

Kanada telah berinvestasi dalam jaringan informasi dan komunikasi dalam bentuk infrastruktur teknis dan layanan masyarakat, seperti perpustakaan dan lembaga layanan sosial. Investasi ini akan terus menambah nilai uang bagi warga Kanada dengan memperluas teknologi Web dari media komunikasi satu arah ke lingkungan kolaboratif.

# KEUNTUNGAN OPEN DATA

**Dukungan untuk penelitian** - Akses ke federasi data pendukung penelitian berbasis penelitian primer di akademisi Kanada dan internasional, sektor publik, dan komunitas penelitian berbasis industri. Akses ke koleksi data, laporan, publikasi, dan artefak yang disimpan di lembaga federal memungkinkan penggunaan koleksi ini oleh para peneliti

**mendukung keputusan yang terinformasi bagi konsumen** - Menyediakan akses ke informasi layanan sektor publik untuk mendukung pengambilan keputusan yang terinformasi; misalnya, statistik perjalanan udara realtime dapat membantu pelancong memilih maskapai penerbangan dan memahami faktor-faktor yang dapat menyebabkan penundaan penerbangan

# KEUNTUNGAN OPEN DATA

**Disclosure Proaktif** – secara proaktif menyediakan data yang relevan dengan warga dan mengurangi jumlah akses ke permintaan informasi, kampanye email, dan pertanyaan media. Ini sangat mengurangi biaya administrasi dan beban yang terkait dengan menanggapi pertanyaan semacam itu

# Open Government Analytics

## Open Government Analytics

adalah cara untuk melacak unduhan dataset dan statistik terbuka.

Pengguna juga dapat melihat pengunjung serta peningkatan partisipasi departemen dan lembaga Pemerintah dalam menyediakan kumpulan data yang lebih terbuka.



# Open Government Analytics

## Open Government Analytics

adalah cara untuk melacak unduhan dataset dan statistik terbuka.

Pengguna juga dapat melihat pengunjung serta peningkatan partisipasi departemen dan lembaga Pemerintah dalam menyediakan kumpulan data yang lebih terbuka.





# Big Data

## BIG DATA

adalah data yang banyak.

Big data adalah jumlah data yang **tidak dapat ditangani dengan menggunakan metode tradisional** dan sangat relatif karena big data lima tahun yang lalu bukanlah data besar hari ini sehingga terus berkembang.

Setiap kali kapasitas penyimpanan seseorang, individu atau perusahaan atau kemampuan untuk menganalisis datanya melebihi jumlah data yang mereka miliki, itu adalah Big Data bagi mereka. Jika kita memiliki banyak data, kita dapat melakukan visualisasi, kita dapat melakukan analisa

# Big Data

Data Science relevan saat ini karena ada banyak data tersedia. Dulu khawatir kekurangan data, sekarang banjir data.

Di masa lalu, belum ada algoritmanya, sekarang banyak algoritma

Dulu software harganya mahal, sekarang open source dan gratis

Dulu, tidak bisa menyimpan data dalam jumlah besar, sekarang dapat memiliki trilyunan set data dengan biaya yang sangat rendah.

Jadi alat untuk bekerja dengan data, ketersediaan data dan kemampuan untuk menyimpan dan menganalisis data, semuanya murah, semuanya tersedia, semuanya ada di mana-mana, ada di sini,

# BIG DATA

Ide dasar di balik ungkapan Big Data adalah bahwa semua yang kita lakukan semakin meninggalkan jejak digital yang dapat kita **gunakan dan analisis untuk menjadi lebih pintar**. Kekuatan pendorong di dunia baru yang berani ini adalah akses ke volume data yang terus meningkat dan kemampuan teknologi yang terus meningkat untuk menambang data tersebut untuk insight komersial.

## **Bernard Marr**

mendefinisikan Big Data sebagai jejak digital yang dihasilkan di era digital ini. Jejak digital ini terdiri dari semua data yang ditangkap saat kita menggunakan teknologi digital.

## **Firma riset Gartner,**

Big Data adalah aset informasi bervolume tinggi, berkecepatan tinggi, dan/atau beraneka ragam yang menuntut bentuk pemrosesan informasi yang hemat biaya dan inovatif yang memungkinkan peningkatan wawasan, pengambilan keputusan, dan proses otomatisasi

## Ernst and Young

Big Data mengacu pada volume data yang dinamis, besar, dan berbeda yang dibuat oleh orang, alat, dan mesin. Ini membutuhkan teknologi baru, inovatif, dan skalabel untuk mengumpulkan, menampung, dan memproses secara analitis sejumlah besar data yang dikumpulkan untuk memperoleh insight bisnis realtime yang berhubungan dengan konsumen, risiko, laba, kinerja, manajemen produktivitas, dan peningkatan nilai pemegang saham.

## Lisa Arthur, kontributor Forbes

mendefinisikan Big Data sebagai kumpulan data dari sumber tradisional dan digital di dalam dan di luar perusahaan yang mewakili sumber penemuan dan analisis berkelanjutan. Tidak ada satu definisi pun tentang Big Data, tetapi ada elemen tertentu yang umum di berbagai definisi yang berbeda seperti **kecepatan, volume, variasi, dan kebenaran.**

# BIG DATA

## FIVE V'S

These are the V's  
of Big Data

**Velocity** adalah kecepatan data, atau kecepatan di mana data terakumulasi.

**Volume** adalah skala data atau pertambahan jumlah data yang disimpan.

**Varietas** adalah keragaman data. Kami memiliki data terstruktur yang cocok dengan rapi ke dalam baris dan kolom, atau database relasional dan data tidak terstruktur yang tidak diatur dengan cara yang telah ditentukan sebelumnya, misalnya Tweet, posting blog, gambar, angka, dan bahkan data video.

**Veracity** adalah kesesuaian dengan fakta dan akurasi.



# BIG DATA

## FIVE V'S

Let's look at some examples of the V's in action.

**Velocity** : Setiap 60 detik, banyak rekaman berjam-jam diunggah ke YouTube. Jumlah data ini dihasilkan setiap menit. Jadi pikirkan tentang berapa banyak yang terakumulasi selama berjam-jam, sehari-hari, dan bertahun-tahun.

**Volume**: Setiap hari terbuat sekitar 2,5 triliun byte data. Itu = 10 juta DVD Blu-ray setiap hari. Populasi dunia adalah sekitar tujuh miliar orang, dan sebagian besar orang sekarang menggunakan perangkat digital



# BIG DATA

## FIVE V'S

Let's look at  
some examples  
of the V's in  
action.

**Variety:** Mari kita pikirkan tentang berbagai jenis data, teks, gambar, dan film. Bagaimana dengan suara, data kesehatan dari perangkat yang dapat dikenakan, dan berbagai jenis data dari perangkat yang terhubung ke internet of things.

**Veracity** : 80% data dianggap tidak terstruktur dan kita harus menemukan cara untuk menghasilkan insight yang handal dan akurat. Data harus dikategorikan, dianalisis dan divisualisasikan.

# BIG DATA

## FIVE V'S

Let's look at some examples of the V's in action.

V terakhir mengacu pada kemampuan dan kebutuhan kita untuk mengubah data menjadi **Value (Nilai)**. Nilai bukan hanya keuntungan. Nilai dapat berupa manfaat medis atau sosial, atau pelanggan, karyawan, atau kepuasan pribadi. Alasan utama mengapa orang menginvestasikan waktu untuk memahami Big Data adalah untuk mendapatkan nilai darinya.

Banyak dari kita menghasilkan dan menggunakan big data tanpa menyadarinya. Bagaimana big data memengaruhi bisnis dan orang-orang?

Pernahkah Anda mencari atau membeli produk di Marketplace? Apakah Anda memperhatikan bahwa Marketplace mulai membuat rekomendasi?

Terkait dengan produk yang Anda cari? Mesin rekomendasi adalah aplikasi umum dari big data.

# Big Data in Business

Perusahaan seperti Tokopedia, Netflix, dan Spotify menggunakan algoritme berdasarkan big data untuk membuat rekomendasi spesifik berdasarkan preferensi pelanggan dan perilaku historis

Asisten pribadi seperti Siri di perangkat

Apple menggunakan big data untuk menyusun jawaban atas pertanyaan yang tak terbatas yang mungkin diajukan pengguna.

Google sekarang membuat rekomendasi berdasarkan big data di perangkat pengguna.

# Big Data in Business

# Big Data in Business

Bagaimana sebuah perusahaan mendapatkan keunggulan kompetitif? Pernahkah Anda mendengar acara Netflix berjudul House of Cards?

Musim pertama acara ini dirilis pada tahun 2013 dan langsung menjadi hit. Pada saat itu, New York Times melaporkan bahwa Eksekutif Netflix tahu bahwa House of Cards akan menjadi hit bahkan sebelum mereka memfilmkannya, tetapi bagaimana mereka tahu itu? Big Data

# Big Data in Business

Netflix memiliki banyak data. Netflix mengetahui waktu saat film ditonton. Mencatat saat pengguna menjeda, memundurkan, dan memajukan dengan cepat. Netflix memiliki peringkat dari jutaan pengguna serta informasi tentang pencarian yang mereka lakukan.

Dengan melihat semua big data ini, Netflix tahu banyak penggunanya telah melakukan streaming karya David Fincher dan film-film yang menampilkan Kevin Spacey selalu berhasil. Dan ia tahu bahwa House of Cards versi Inggris juga berhasil dengan baik.

# **BIG DATA & DATA SCIENCE**

Mari kita lihat contoh Walmart.

Walmart menggunakan mesin pencari yang disebut Polaris, yang membantu pembeli mencari produk yang ingin mereka beli.

Ini memperhitungkan bagaimana perilaku pengguna di situs web untuk menampilkan hasil terbaik bagi mereka.

Polaris akan memunculkan item yang didasarkan pada minat pengguna dan, karena banyak konsumen mengunjungi situs web Walmart, sejumlah besar data dikumpulkan, membuat analisis terhadap big data tersebut menjadi sangat penting.



# BIG DATA & DATA SCIENCE

Data Scientist adalah orang yang memenuhi syarat untuk memperoleh insight dari data dengan menggunakan keterampilan dan pengalaman dari ilmu komputer, bisnis, atau sains, dan statistik.

Ada lebih banyak keterampilan yang harus dimiliki seorang Data Scientist. Seseorang dapat menggunakan proses berikut untuk memahami data besar.

**Determine Problem.** Apa masalah bisnisnya? Apa tujuan proyek? Apa yang akan Anda lakukan jika Anda memiliki semua data?

# Big Data and Data Science Key Steps

PENGUMPULAN DATA. DATA  
MANA YANG RELEVAN? APAKAH  
ADA MASALAH PRIVASI?

- Eksplorasi datanya
- Plot datanya. Apakah ada pola?
- Analisis data
- Membangun model
- Sesuaikan modelnya
- Validasi model
- Bercerita
- Visualisasi plus komunikasi
- Can we tell a story? Take action and make decisions

# Summary

- Data scientist is to be **curious, extremely argumentative, judgmental.**

- The last thing that a data scientist needs and that is the ability to **tell a story.**

- **Because if you don't tell a great story from it, your findings will remain hidden, it will remain buried, nobody would know, but your rise to prominence is pretty much relying on your ability to tell great stories.**

# **Apa yang anda dapatkan dari kuliah hari ini**

**BAGIKAN PENDAPAT ANDA DENGAN REKAN  
SEKELAS.**