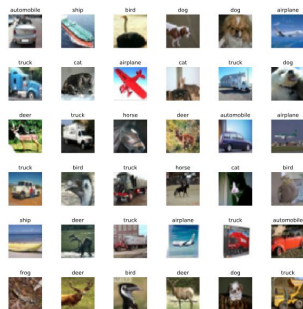


Data Analytics

Mrinal Das

Types of Analytics

Data



Hillary Clinton ✓

@HillaryClinton

Following

"To all the little girls watching...never doubt that you are valuable and powerful & deserving of every chance & opportunity in the world."

4:51 PM - 9 Nov 2016

675,019 Retweets 1,253,910 Likes



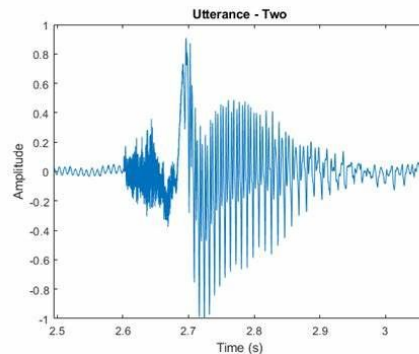
47K

675K

1.3M



Team	Span	Mat	Won▼	Lost	Tied	Draw	W/L	Ave	RPO
Australia	1877-2024	866	414	232	2	218	1.784	34.39	2.97
England	1877-2024	1071	392	324	0	355	1.209	32.19	2.78
West Indies	1928-2024	575	183	210	1	181	0.871	31.46	2.98
India	1932-2024	579	178	178	1	222	1.000	33.79	2.95
South Africa	1889-2024	464	178	161	0	125	1.105	31.56	2.82
Pakistan	1952-2024	456	148	142	0	166	1.042	32.67	2.92
New Zealand	1930-2024	470	115	185	0	170	0.621	29.58	2.71
Sri Lanka	1962-2024	316	103	121	0	92	0.851	32.50	3.11
Bangladesh	2000-2024	142	19	105	0	18	0.180	25.07	3.14
Zimbabwe	1992-2023	117	13	75	0	29	0.173	26.14	2.71
Afghanistan	2018-2024	9	3	6	0	0	0.500	22.69	3.03
Ireland	2018-2024	8	1	7	0	0	0.142	23.48	3.03
ICC World XI	2005-2005	1	0	1	0	0	0.000	16.70	3.43



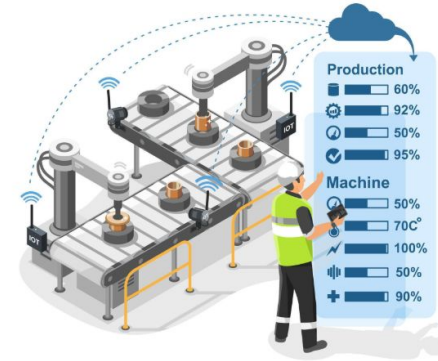
Goal of Data Analytics

Scientific ways to store, process, extract data

Data holds history, but hard to understand

Analysing data can reveal

- Cause of failure
- Possible future
- Ways to make future better



Case Study

You have a company to transport eggs to various towns.

You just checked the journal and observed:

10 July	2000
11 July	1800
12 July	1500
13 July	1400

Case Study

You have a company to transport eggs to various towns.

You just checked the journal and observed:

10 July	2000
11 July	1800
12 July	1500
13 July	1400

1. What will be the sale for 14 July ?
2. Is there any issue ?
3. What is the issue ?
4. What will you do ?

Example

Lamborghini manufactures **10,000 cars every month**. Each car passes through several stages:

- Body assembly
- Painting
- Engine installation
- Quality inspection

The company wants to increase profit, reduce defects, and improve customer satisfaction.

Example

Physical process:

Manufacturing cars

- Body assembly
- Painting
- Engine installation
- Quality inspection
- Delivery to dealers

Goal: reduce defects

Possible failures:

- Body parts defective
- Painting unfinished
- Wrong installation

Data Associated:

- Various sensor signals
- Physical observations
- Images and quality reports

Example

Goal: reduce defects

Important steps:

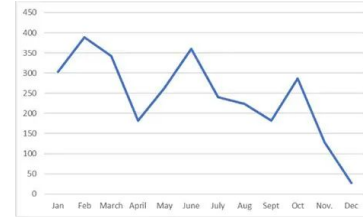
1. Identify the issue
2. Investigate the cause
3. Foresee other issues
4. Devise a solution

Example

Goal: reduce defects

Important steps:

1. Identify the issue
2. Investigate the cause
3. Foresee other issues
4. Devise a solution



Example

Goal: reduce defects

Important steps:

1. Identify the issue
2. Investigate the cause
3. Foresee other issues
4. Devise a solution



Example

Goal: reduce defects

Important steps:

1. Identify the issue
2. Investigate the cause
3. Foresee other issues
4. Devise a solution



Example

Goal: reduce defects

Important steps:

1. Identify the issue
2. Investigate the cause
3. Foresee other issues
4. Devise a solution



Use of Analytics

Descriptive

What happened

Predictive

What will
happen

Diagnostic

Why happened

Prescriptive

What to do

Descriptive Analytics (What happened?)

Scenario

At the end of the month,
managers review production data.

Metric	Value
Cars produced	10,000
Defective cars	500
Average assembly time	8 hours
Machine downtime	120 hours
Warranty claims	200

Insights

- Defect rate = 5%
- Production was lower than expected.

Business Question

"What happened in the factory last month?"

Descriptive Analytics

Data is huge for someone to manually read through it.

Processing and extracting pattern is nearly impossible.

We need some scientific method to summarize for easy comprehension.

1071	392	324	0	355	1.209	32.19	2.78
575	183	210	1	181	0.871	31.46	2.98
579	178	178	1	222	1.000	33.79	2.95
464	178	161	0	125	1.105	31.56	2.82
456	148	142	0	166	1.042	32.67	2.92
470	115	185	0	170	0.621	29.58	2.71
316	103	121	0	92	0.851	32.50	3.11
142	19	105	0	18	0.180	25.07	3.14
117	13	75	0	29	0.173	26.14	2.71
9	3	6	0	0	0.500	22.69	3.03
8	1	7	0	0	0.142	23.48	3.03
1	0	1	0	0	0.000	16.70	3.43

Descriptive Analytics

Statistics can provide information automatically.

Why do we need to learn data analytics then ?

- ★ A data analyst can decide which statistic to check.
- ★ Often, it may not be straightforward, and one needs to try several statistics.
- ★ A data analyst can connect one information to another.
- ★ Analyst can identify the key performance indicators (KPI) and measure them.

Methods of Descriptive Analytics

Method	Purpose
Summary Statistics	Describe the center and spread
Frequency Analysis	Count occurrences
Cross-tabulation	Compare categories
Data Aggregation	Summarize detailed data
Data Visualization	Present information graphically
Trend Analysis	Study changes over time
Distribution Analysis	Understand data shape
Percentages and Ratios	Measure proportions
Ranking and Sorting	Compare performance

Diagnostic Analytics (Why did it happen?)

Descriptive analytics can reveal some anomaly or abnormality

Diagnostic analytics can be used to fix the issue

To fix the issue one needs to identify the cause of the issue

Method of Diagnostic Analytics

Step 1: Observe the pattern in descriptive analytics

Step 2: Drill down by slicing and grouping

Step 3: Find out correlation between condition and effect

Methods of Diagnostic Analytics

Drill down analysis

Example: overall defect rate 5%

Further drill down plant B at department level

Drill down at plant level

Plant	Defect Rate
Plant A	2%
Plant B	7%
Plant C	3%

Department	Defect Rate
Assembly	2%
Paint Shop	10%
Inspection	3%

Insight: something is wrong at the paint shop

Analyse further data of paint shop in plant B

Temperature	Humidity	Paint Defect
22	45	No
30	85	Yes
31	88	Yes
24	50	No

Paint defects increase when humidity exceeds 80%.

Diagnostic Analytics (Why did it happen?)

Scenario

Engineers investigate why paint defects increased.

Data Collected

Temperature (°C)	Humidity (%)	Paint Defect
22	45	No
30	85	Yes
31	88	Yes
24	50	No

Analysis

Correlation and statistical analysis reveal:

- Paint defects increase when humidity exceeds 80%.
- A malfunctioning air-conditioning unit caused excessive humidity.

Business Question

"Why are so many cars failing paint inspection?"

Outcome

Maintenance fixes the climate control system.

Defect rate decreases from **5% to 2%**.

Predictive Analytics (What is likely to happen?)

Scenario

Sensors monitor robotic welding machines.

Measurements include:

- Temperature
- Vibration
- Pressure
- Running hours

Historical data show that machines usually fail before vibration reaches a critical threshold.

Methods of Predictive Analytics

Using historical data we predict the future

Machine learning, Deep learning are common techniques

Method	Prediction
Regression	Remaining useful life (hours)
Classification	Will the machine fail? (Yes/No)
Time Series	Next week's vibration level
Decision Tree	Identify failure conditions
Random Forest	Predict failures with high accuracy
Neural Network	Learn complex sensor patterns
Survival Analysis	Estimate time until failure

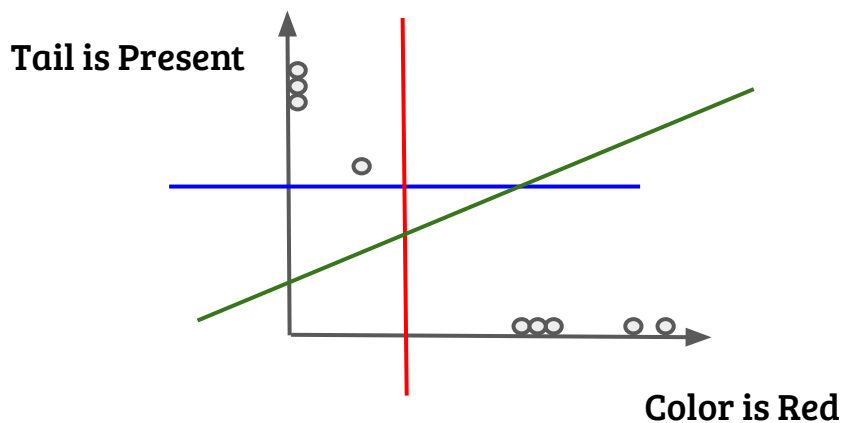
Data for AI



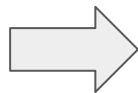
VS



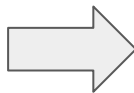
- Collect samples
- Plot points
- Find a boundary to separate classes.
- The Separating Line is the Classifier.



Information Extraction from Data

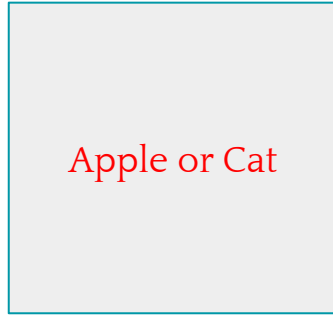


Features	Color=Red	Shape=round
Statistics	70%	100%



Features	Tail=Yes	Shape=round
Statistics	100%	0%

Prediction using ML

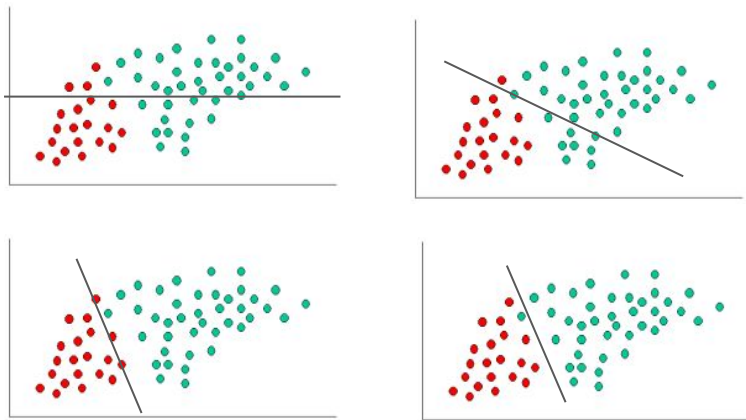


Features	Color=Red	Shape=round
Observation	Yes	Yes

Features	Tail	Shape=round
Observation	No	Yes

Task: write a pseudo-code for the above logic

Machine Learning with Data



ML can be interpreted as searching.

A suitable function to achieve the goal.

Example:

A line has three variables.

Search values for those variables

So that it separates two groups (classes).

Predictive Analytics (What is likely to happen?)

Machine	Probability of Failure Next Week
Robot A	10%
Robot B	90%
Robot C	20%

Business Question

"Which machines are likely to fail in the near future?"

Outcome

Maintenance is performed before Robot B breaks down.

Analysis helps to avoid:

- Production delays
- Expensive repairs
- Lost revenue

This is called **predictive maintenance**.

Prescriptive Analytics (What should we do?)

Prescriptive analytics is the most advanced form of data analytics. It goes beyond predicting what is likely to happen by **recommending the best course of action** to achieve a desired objective.

Prescriptive analytics uses mathematical optimization, simulation, business rules, and artificial intelligence to recommend actions that maximize benefits or minimize costs while satisfying real-world constraints.

Prescriptive Analytics

Example

A car manufacturer predicts that demand for SUVs will increase by 20% next month (**predictive analytics**).

Prescriptive analytics answers:

- How many SUVs should be produced?
- Which factory should manufacture them?
- How should workers and machines be scheduled?
- How can production costs be minimized while meeting demand?

Predictive vs Prescriptive

SUV demand is expected to increase by 20%.

Increase SUV production by 15%.

Robot R12 has an 85% chance of failure next week.

Schedule maintenance for Robot R12 this weekend.

Battery inventory will run out in 10 days.

Order 5,000 additional batteries today.

Applications of Prescriptive Analytics

- ★ Production planning
- ★ Transportation
- ★ Supply chain
- ★ Resource allocation
- ★ Scheduling workers
- ★ Assigning machines
- ★ Vehicle routing

Methods of Prescriptive Analytics

Optimization: find out the best solution given the resources and constraints

Simulation: Simulate different scenarios and evaluate outcomes

Rule based: Derive rules based on experience and use

Decision Tree: Build decision tree based on past and use it to evaluate outcomes

Reinforcement learning: learn the best decision by interacting with environment

Markov decision process: make decisions sequentially

Digital twin: receive data from physical environment and validate decisions

Prescriptive Analytics (What should we do?)

Scenario

Suppose demand forecasts predict:

- SUVs: 4,000 orders
- Sedans: 2,500 orders
- Electric cars: 3,500 orders

The company has limited:

- Workers
- Batteries
- Assembly lines

Optimization algorithms recommend:

- Assign Line 1 to electric vehicles.
- Increase SUV production by 15%.
- Schedule maintenance during weekends.
- Order additional batteries immediately.

Business Question

"What actions should we take to maximize profit?"

Outcome

The company:

- Increases profit by 12%.
- Reduces delivery delays.
- Uses resources more efficiently.

Summary

Type of Analytics	Main Question	Car Manufacturing Example
Descriptive	What happened?	Paint defect rate increased to 5%
Diagnostic	Why did it happen?	High humidity caused paint defects
Predictive	What will happen?	Welding robot likely to fail next week
Prescriptive	What should we do?	Schedule maintenance and optimize production

"Modern car factories generate millions of sensor readings every day; data analytics transforms this raw data into answers about what happened, why it happened, what will happen next, and what actions should be taken to manufacture better cars at lower cost."

Case Study

You have a company to transport eggs to various towns.

You just checked the journal and observed:

10 July	2000
11 July	1800
12 July	1500
13 July	1400

1. What will be the sale for 14 July ?
2. Is there any issue ?
3. What is the issue ?
4. What will you do ?

Exploratory Data Analytics

Suppose a company has collected huge amounts of data from various sources.

But it does not know what to do with that. Or how that data can be useful.

What shall we do ?

Quiz-2

1. What are different analytics ?
2. Take one scenario and explain how different analytics can be useful.
3. Write 2 things that you liked/understood today.
4. Write 2 things that you did not like/understand today.
5. Write a short description of any manufacturing process preferably that you have experience with and scope of data analytics for that.
 - a. Physical process
 - b. Input and output
 - c. Possible failures
 - d. Data associated