Google I/O Extended

Simple ML for Sheets

Simple ML

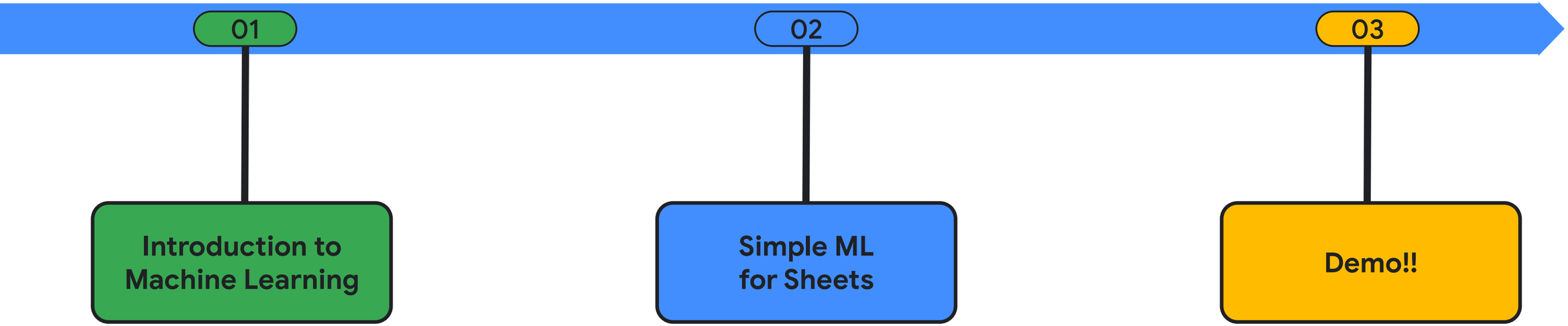
Google I/O Extended

$ whoami
Sandip Palit

- Technology Consultant @ PwC India
- Organizer @ Postman Community Kolkata
- Google Crowdsource Influencer
- Research article on “Real Time Sentiment Analysis”
- Python Library “Relocate”
- Loves to explore new places and new music

Disclaimer: *The thoughts, opinions, and ideas I will be sharing with you today are entirely my own and I am not speaking on behalf of any organization or entity.*

Agenda


01
02
03
Introduction to Machine Learning
Simple ML for Sheets
Demo!!

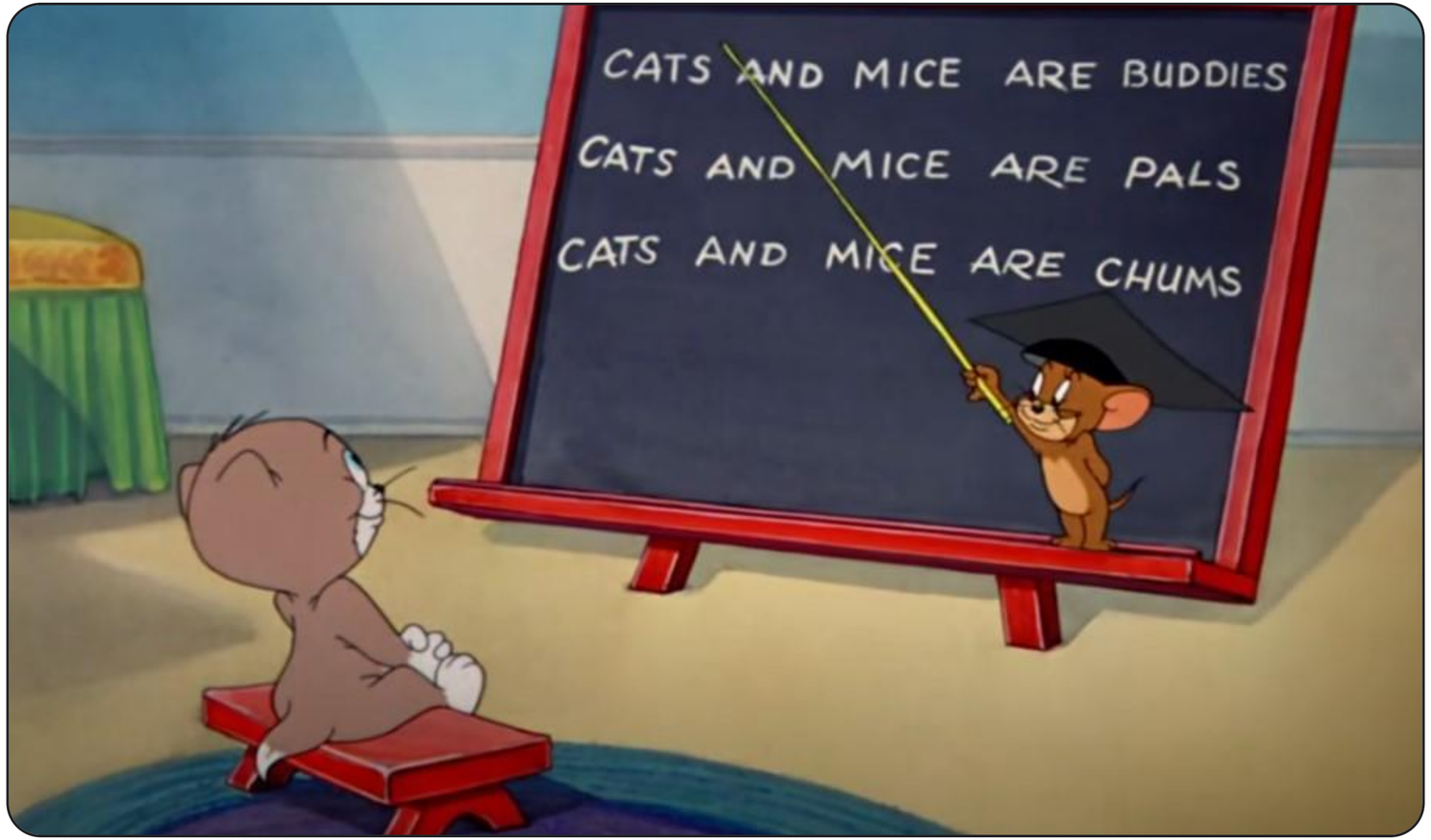
Section 01

Introduction to Machine Learning

Section 01

What is Machine Learning?

- Subset of Artificial Intelligence.
- Enable machines to learn from data.
- Without being explicitly programmed.



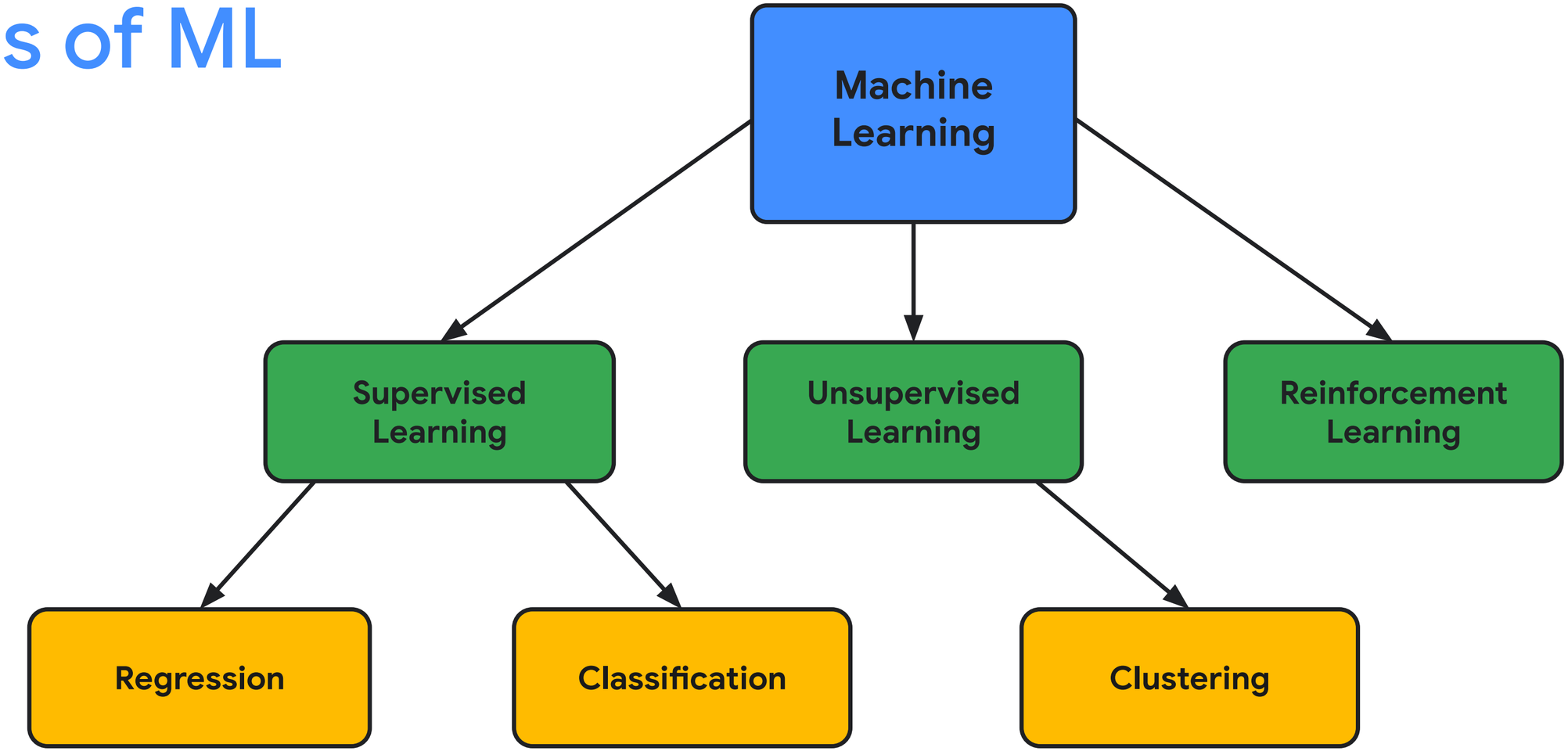
Section 01

Applications of ML

Healthcare	Finance	Retail	Agriculture
Medical Imaging Analysis	Credit Scoring	Demand Forecasting	Crop Yield Prediction
Health Monitoring	Algorithmic Trading	Price Optimization	Disease Detection

Section 01

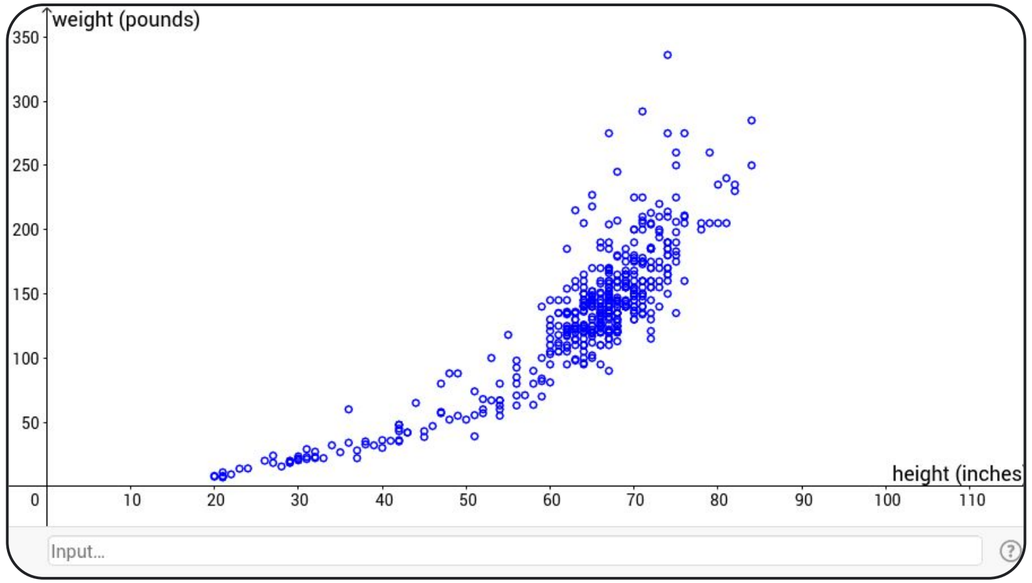
Types of ML



Section 01

Supervised Learning

- Training with labeled data (input-output pairs).
- Goal: Make predictions on new, unseen data.

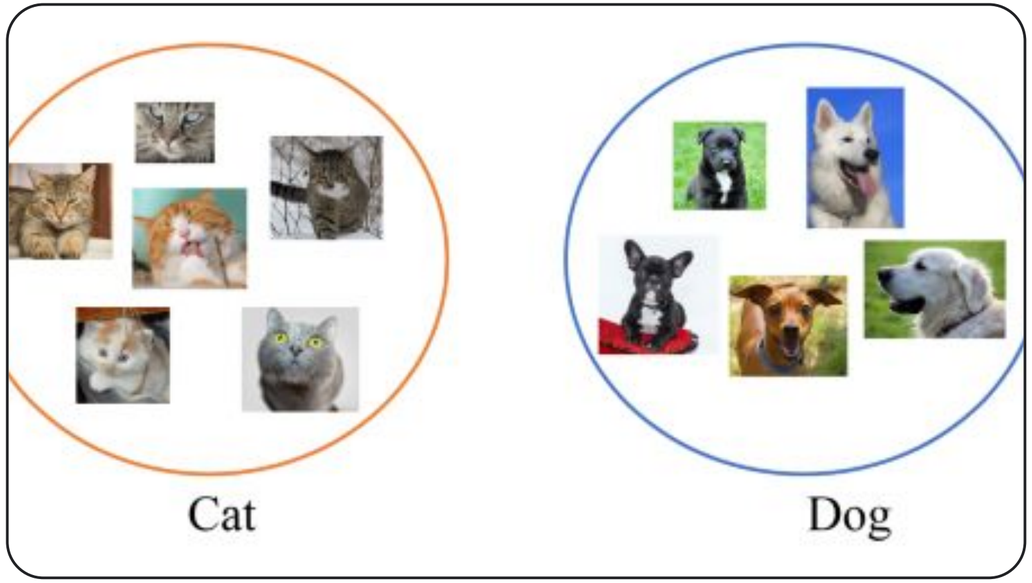


Regression

- On continuous data
- E.g.: Predicting the Weight of a person based on Height.

Classification

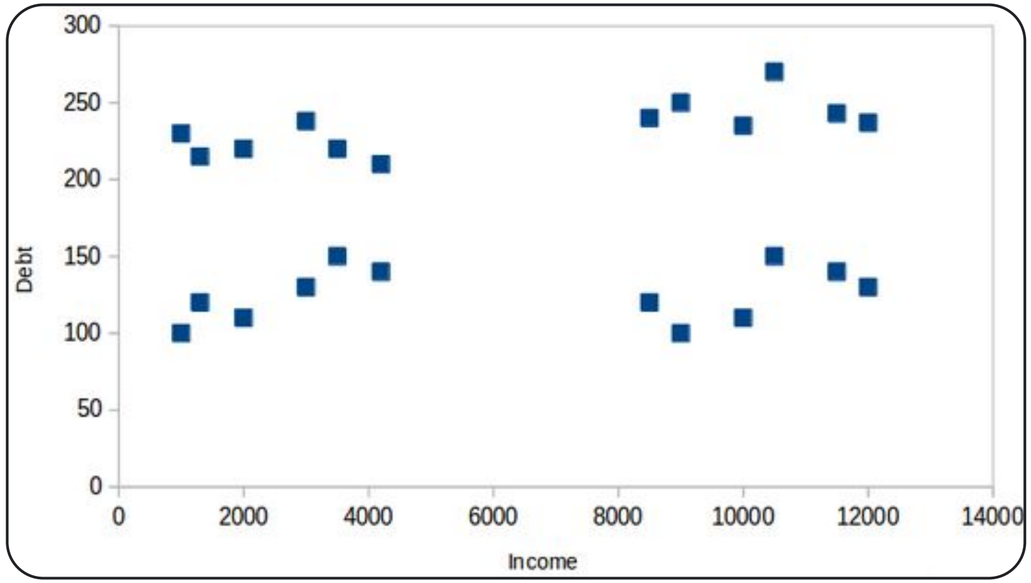
- On categorical data
- E.g.: Predicting whether the image is of a Cat or a Dog.



Section 01

Unsupervised Learning

- No labeled data available.
- Finding patterns in the data.
- Goal: Discover inherent relationships in the data.



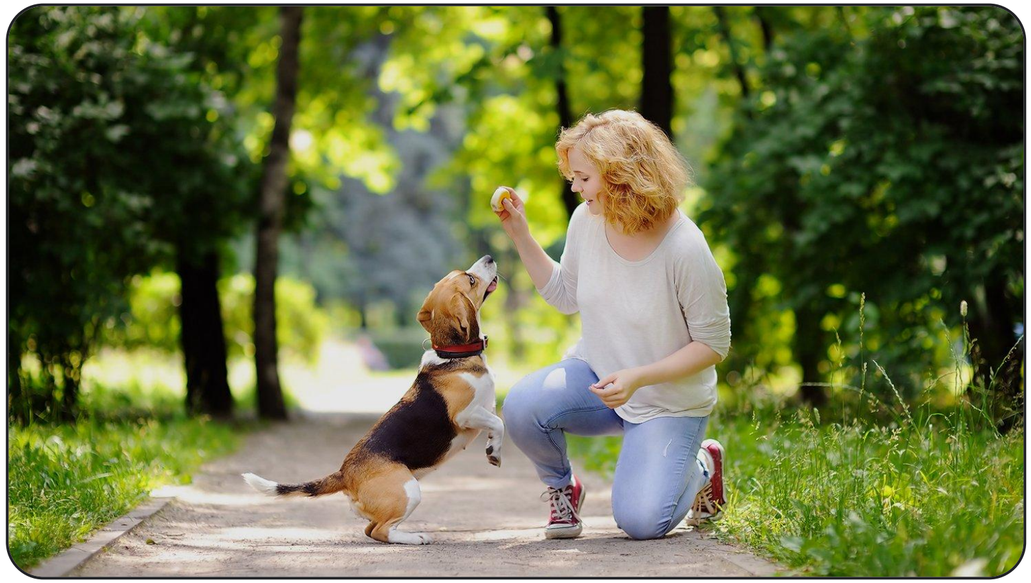
Clustering

- Grouping of similar data points
- E.g.: Grouping individuals based on Debt vs Income.

Section 01

Reinforcement Learning

- Agents learn by interacting with an environment.
- Reward-based learning.
- Goal: Learn optimal actions to maximize rewards.
- E.g.: Training our Pet.



Section 01

Steps of ML

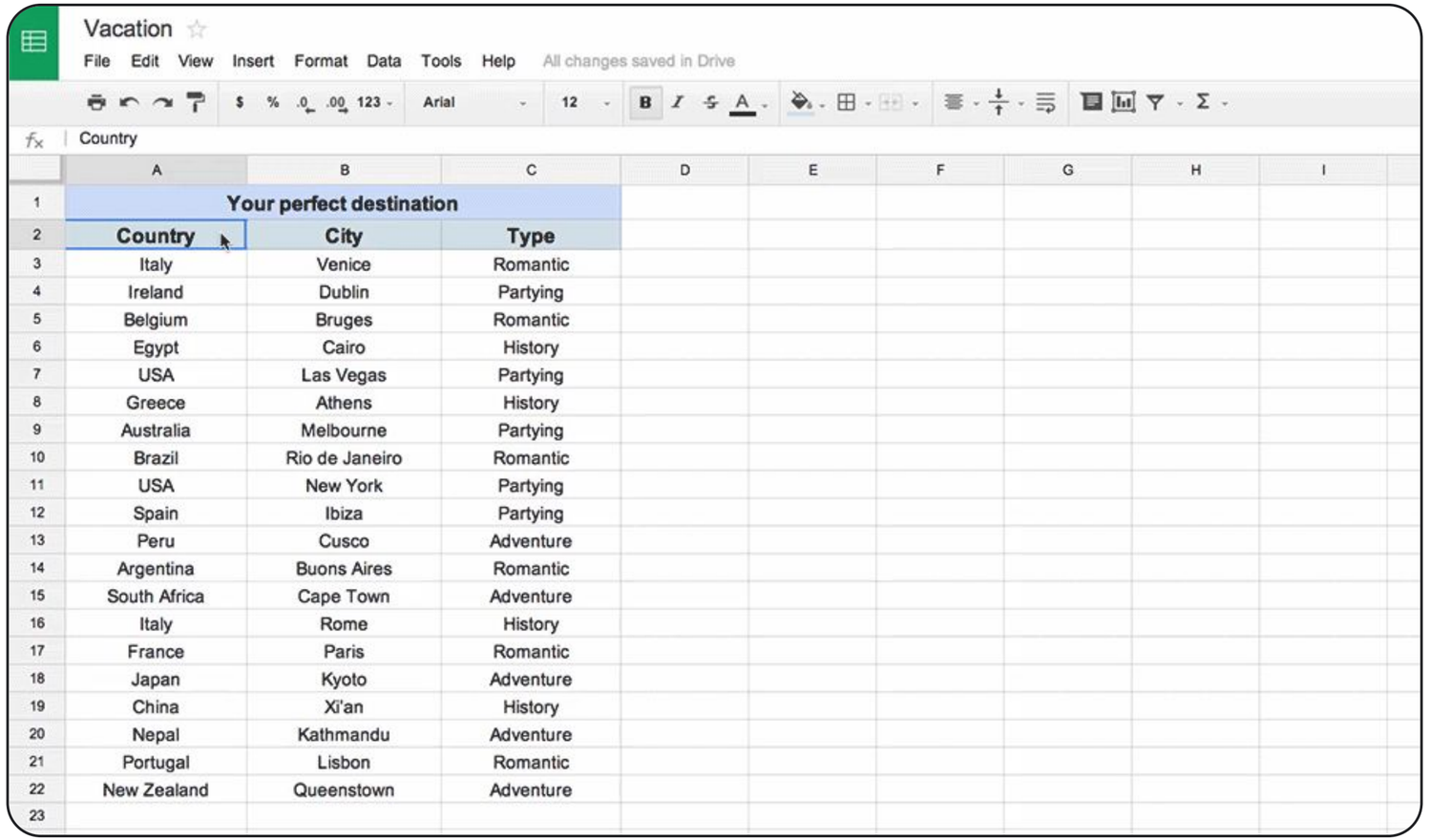
Section 02

Simple ML for Sheets

Section 02

What is Google Sheets?

- It’s a cloud-based spreadsheet application developed by Google.
- Part of the Google Workspace suite of productivity tools.



	A	B	C
1	Your perfect destination		
2	Country	City	Type
3	Italy	Venice	Romantic
4	Ireland	Dublin	Partying
5	Belgium	Bruges	Romantic
6	Egypt	Cairo	History
7	USA	Las Vegas	Partying
8	Greece	Athens	History
9	Australia	Melbourne	Partying
10	Brazil	Rio de Janeiro	Romantic
11	USA	New York	Partying
12	Spain	Ibiza	Partying
13	Peru	Cusco	Adventure
14	Argentina	Buons Aires	Romantic
15	South Africa	Cape Town	Adventure
16	Italy	Rome	History
17	France	Paris	Romantic
18	Japan	Kyoto	Adventure
19	China	Xi'an	History
20	Nepal	Kathmandu	Adventure
21	Portugal	Lisbon	Romantic
22	New Zealand	Queenstown	Adventure
23			

Section 02

Features of Google Sheets

Online Access

Import and Export

Data Visualization

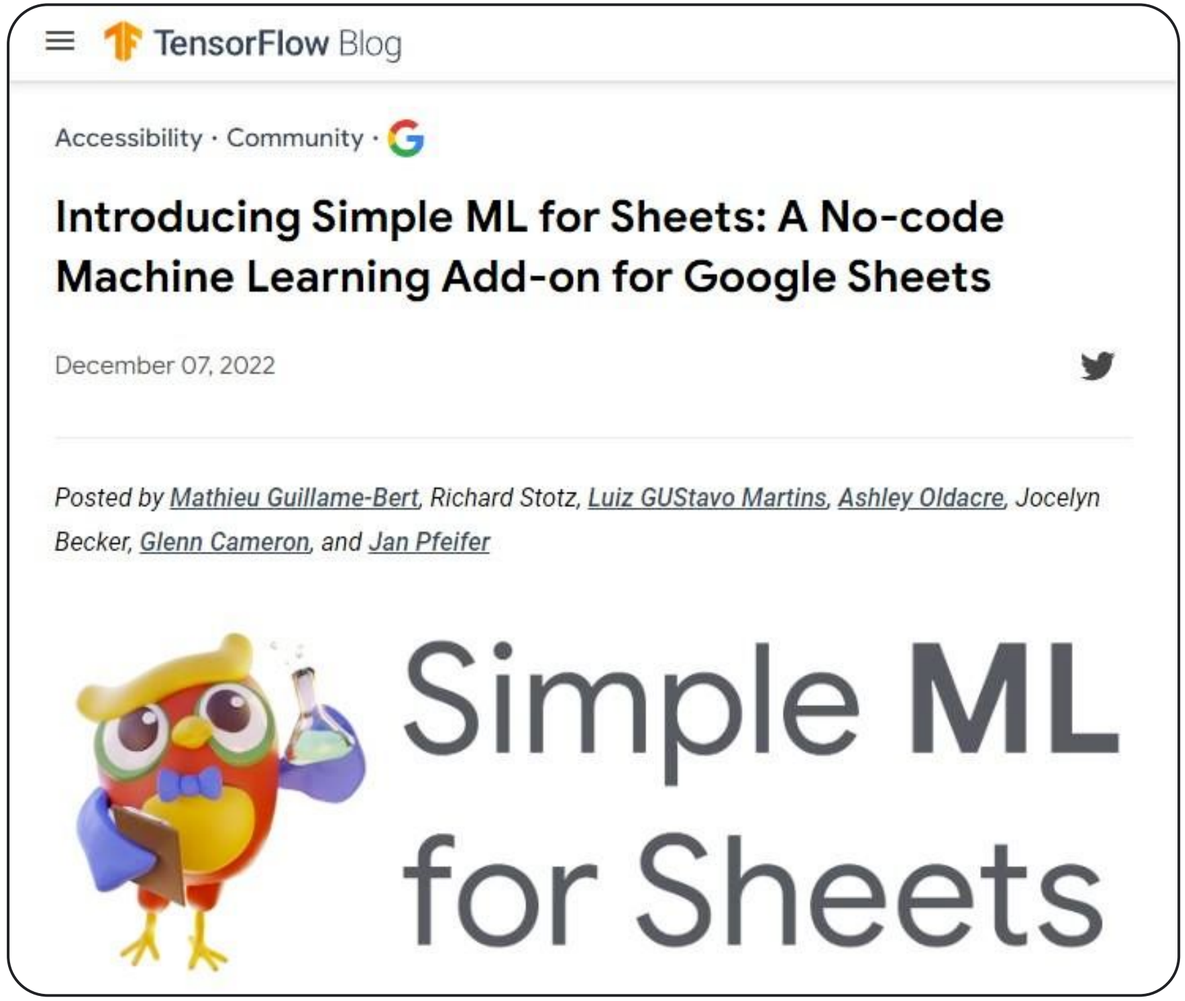
Real-Time Collaboration

Formulas and Functions

Section 02

Simple ML for Sheets

- It's a Google Sheets addon that helps us in using Machine Learning.
- Designed for beginners, it enables us to work without prior coding or ML expertise.
- The model trains locally in our browser within few minutes.
- We can export the models that Simple ML generates to Google Colab.



Section 02

Learning Algorithms

Decision Tree

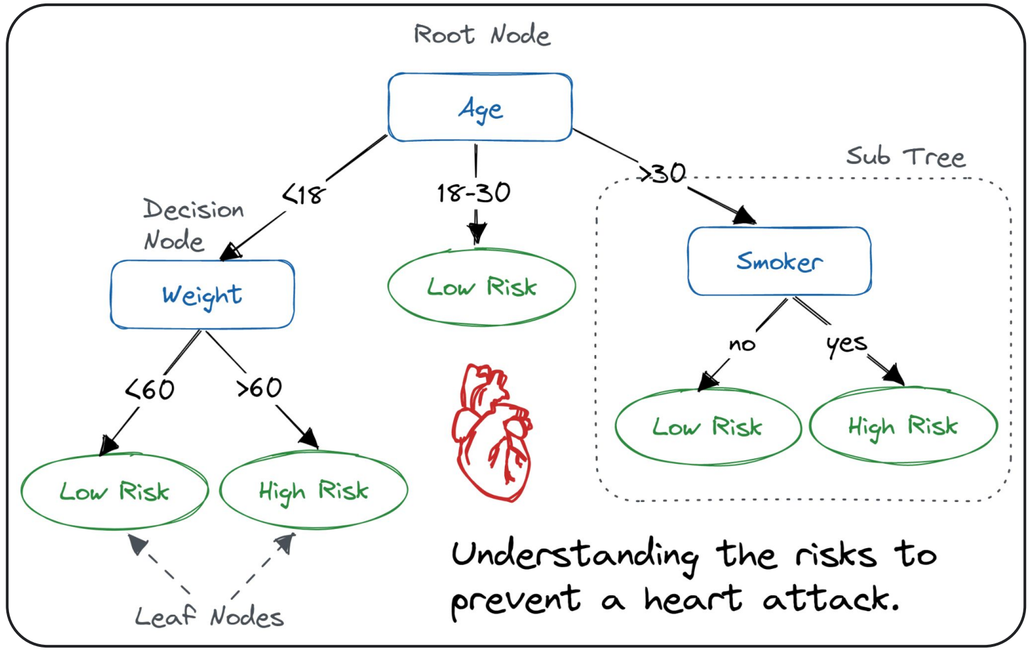
Random Forest

Gradient Boosted Trees

Section 02

Decision Tree

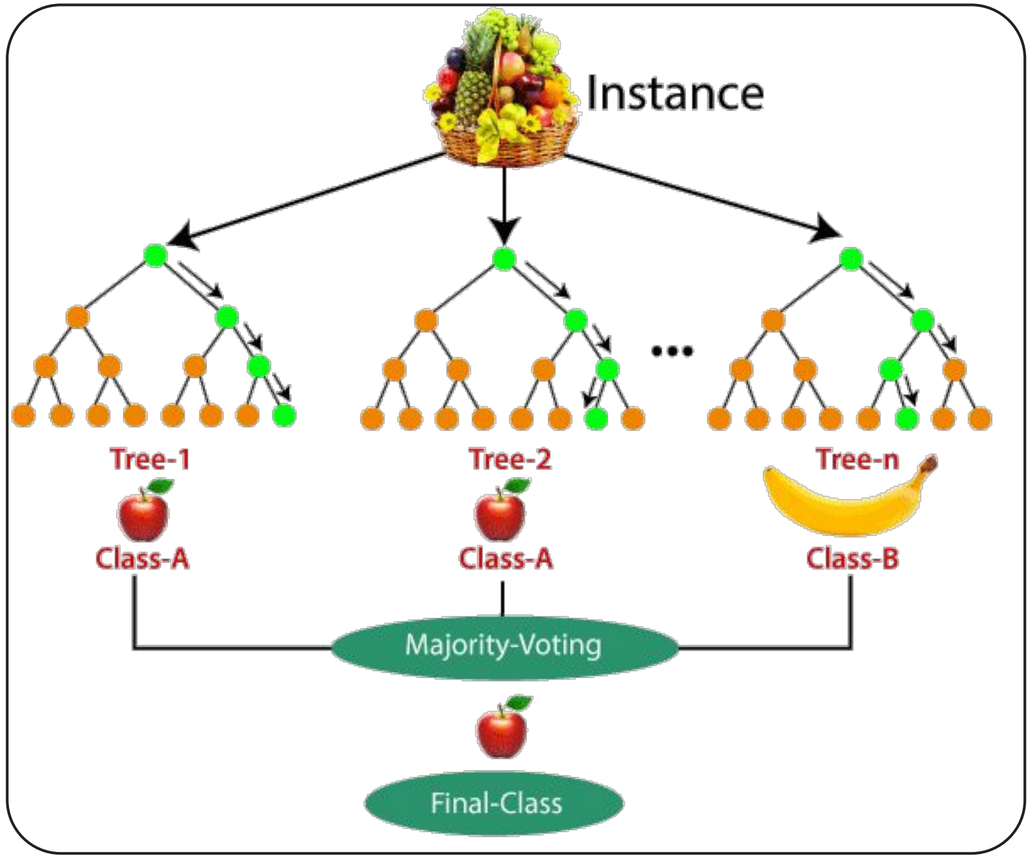
- It's a tree-structured classifier.
- Internal nodes represent the features of the dataset.
- Branches represent the decision rules.
- Each leaf node represents the outcome
- E.g.: Predicting the chances of Heart Attack.



Section 02

Random Forest

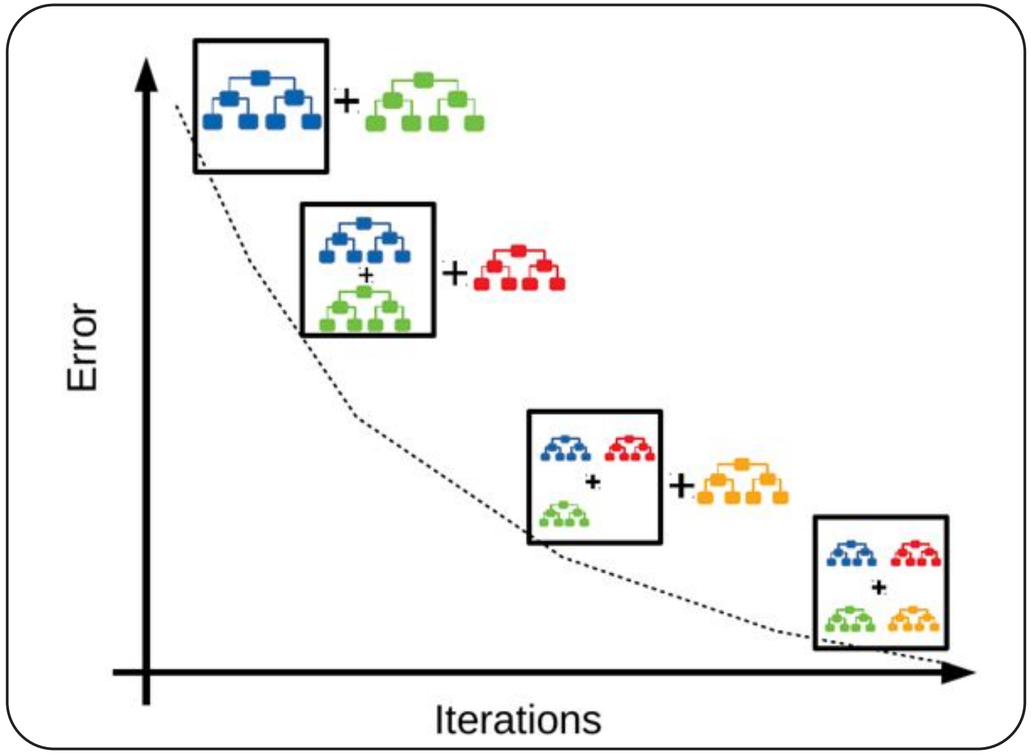
- It’s based on the concept of ensemble learning.
- Contains independent decision trees on subsets.
- Takes the average to improve predictive accuracy.
- Base classifiers are trained parallelly.
- Each model has equal weightage.
- E.g.: Predicting whether it’s an Apple or a Banana



Section 02

Gradient Boosted Trees

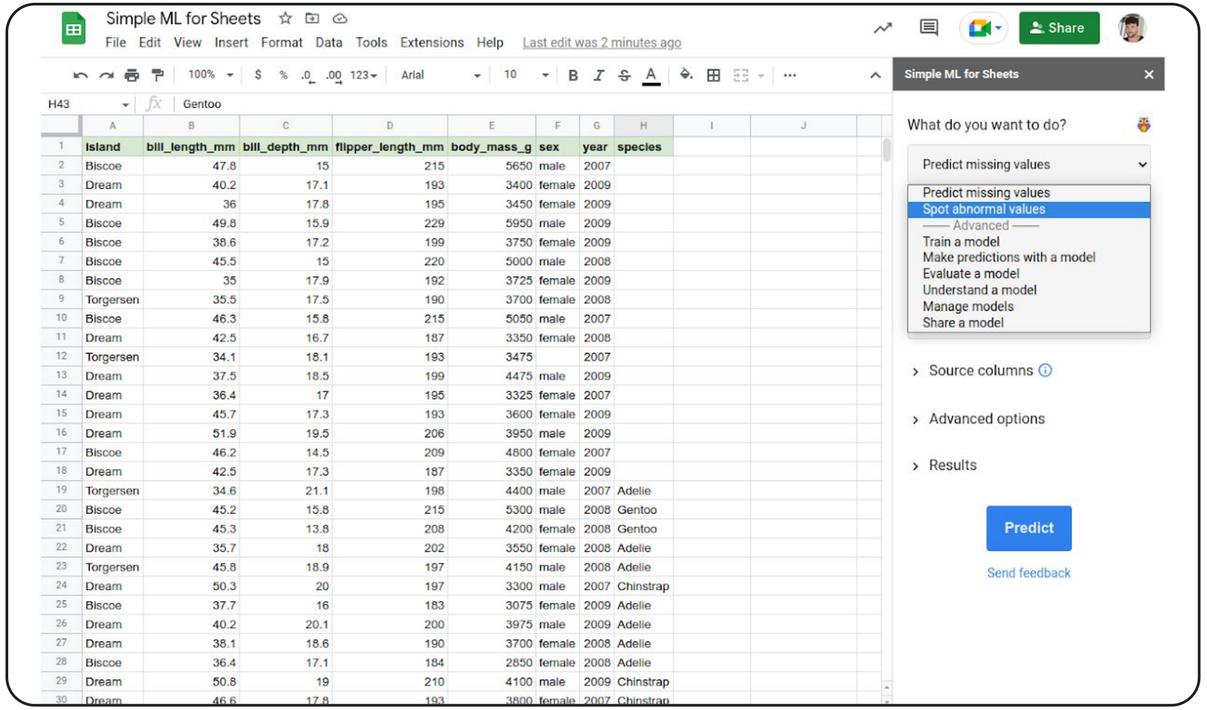
- It's also based on the concept of ensemble learning.
- Firstly, a base model is built from the training data.
- Then the second model is built which tries to correct the errors present in the first model, and so on..
- Base classifiers are trained sequentially.
- Models are weighted according to their performance.



Section 02

Tasks supported by Simple ML

- Predict missing values
- Spot abnormal values
- Forecast future values



	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year	species
2	Biscoe	47.8	15	215	5650	male	2007	
3	Dream	40.2	17.1	193	3400	female	2009	
4	Dream	36	17.8	195	3450	female	2009	
5	Biscoe	49.8	15.9	229	5950	male	2009	
6	Biscoe	38.6	17.2	199	3750	female	2009	
7	Biscoe	45.5	15	220	5000	male	2008	
8	Biscoe	35	17.9	192	3725	female	2009	
9	Torgersen	35.5	17.5	190	3700	female	2008	
10	Biscoe	46.3	15.8	215	5050	male	2007	
11	Dream	42.5	16.7	187	3350	female	2008	
12	Torgersen	34.1	18.1	193	3475		2007	
13	Dream	37.5	18.5	199	4475	male	2009	
14	Dream	36.4	17	195	3325	female	2007	
15	Dream	45.7	17.3	193	3600	female	2009	
16	Dream	51.9	19.5	206	3950	male	2009	
17	Biscoe	46.2	14.5	209	4800	female	2007	
18	Dream	42.5	17.3	187	3350	female	2009	
19	Torgersen	34.6	21.1	198	4400	male	2007	Adelie
20	Biscoe	45.2	15.8	215	5300	male	2008	Gentoo
21	Biscoe	45.3	13.8	208	4200	female	2008	Gentoo
22	Dream	35.7	18	202	3550	female	2008	Adelie
23	Torgersen	45.8	18.9	197	4150	male	2008	Adelie
24	Dream	50.3	20	197	3300	male	2007	Chinstrap
25	Biscoe	37.7	16	183	3075	female	2009	Adelie
26	Dream	40.2	20.1	200	3975	male	2009	Adelie
27	Dream	38.1	18.6	190	3700	female	2008	Adelie
28	Biscoe	36.4	17.1	184	2850	female	2008	Adelie
29	Dream	50.8	19	210	4100	male	2009	Chinstrap
30	Dream	46.6	17.8	193	3800	female	2007	Chinstrap

Section 03

Demo!!

https://bit.ly/sml-sp

Section 03

Use Cases

Predict missing values

Forecast future values

Spot abnormal values

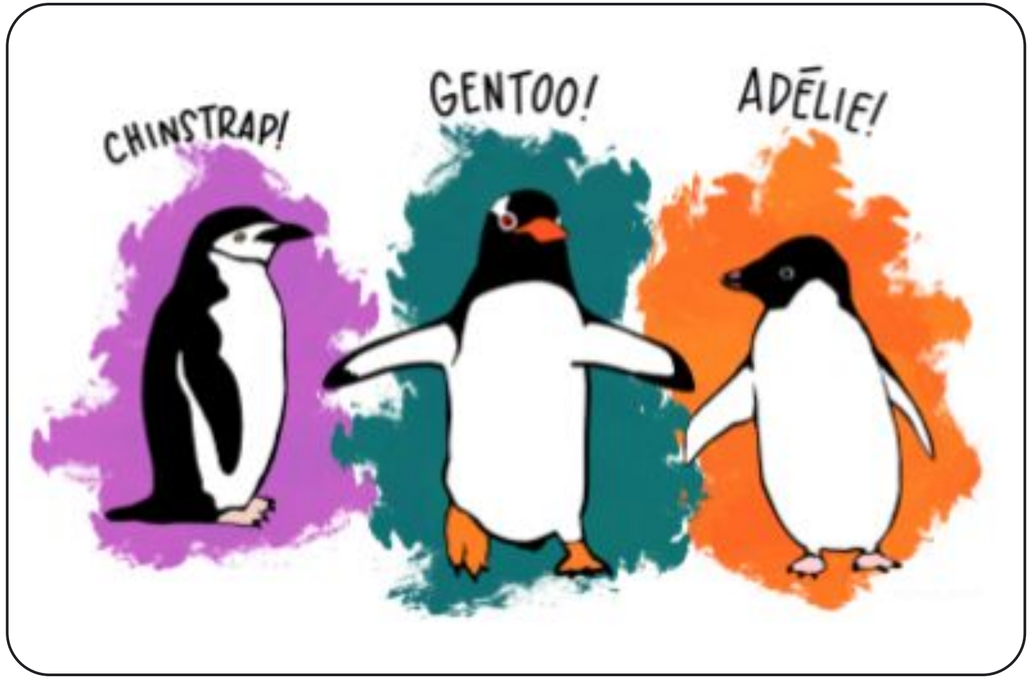
Train & Evaluate Model

Reference: https://simplemlforsheets.com/

Section 03

Predict missing values

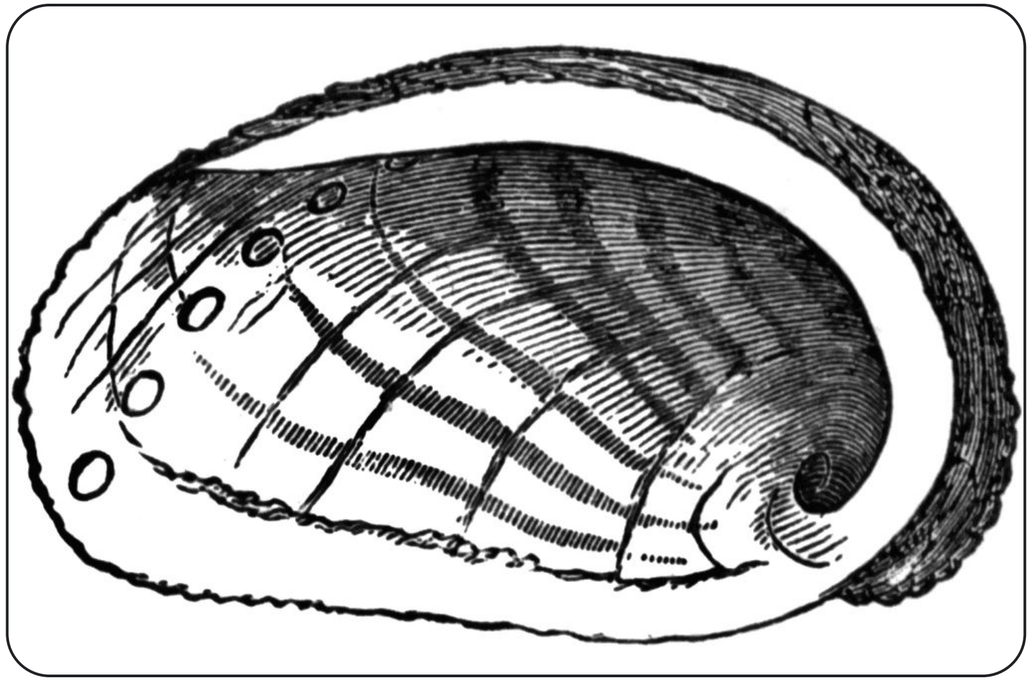
- We are studying a colony of penguins in Antarctica.
- Three penguin species: Chinstrap, Gentoo, and Adelie.
- We collected various measurements of 300 penguins.
- But, we forgot to note the species of 30 penguins.
- We will use Simple ML to recover the species of those 30 penguins.



Section 03

Spot abnormal values

- We are studying the Abalone Snails.
- We collected various measurements of those snails.
- This dataset contains records for around 4000 abalones.
- But there are some human errors in the dataset.
- We will use Simple ML to look for abnormalities in the number of ring.



Section 03

Forecast future values

(On Single Column)

- We are data analysts in an air travel company.
- We have records of the number of monthly passengers from January 1949 to December 1960.
- We will use Simple ML to predict the traffic from January 1961 to May 1963.

Section 03

Forecast future values

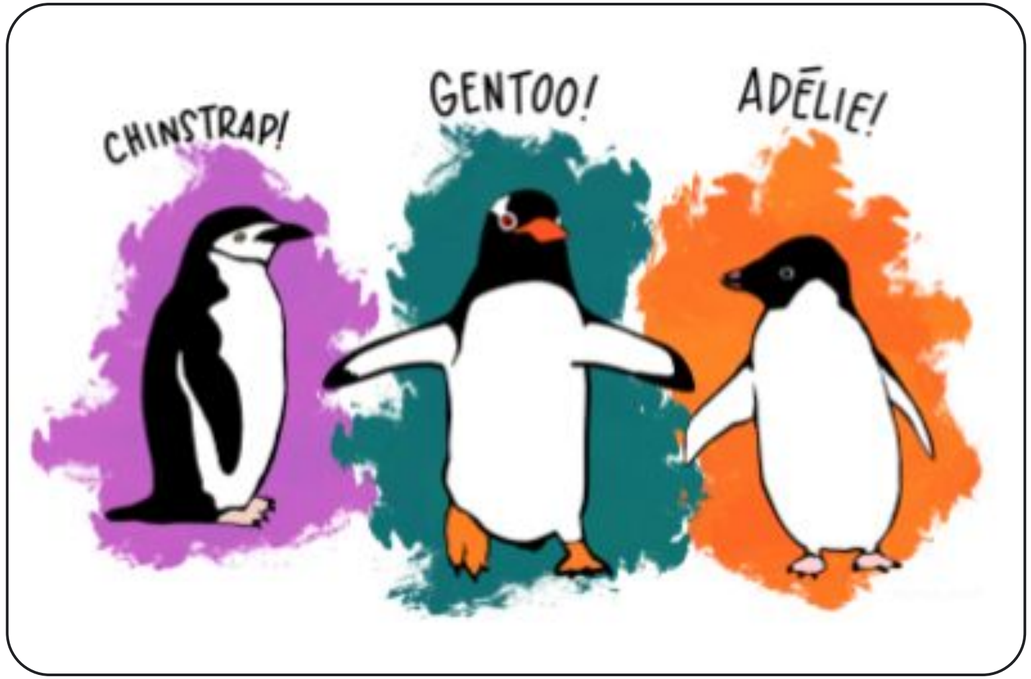
(On Multiple Column)

- We are data analysts hired to analyse the sales of two stores located in London and New York.
- We will use Simple ML to forecast the weekly sales of Toys and Books departments in these shops.

Section 03

Train & Evaluate Model

- We will use the same penguins dataset.
- The dataset is divided into Training set (303 records) and Testing set (42 records).
- Firstly, we will Train a new Model on the Training set.
- Then, we will Understand our trained Model.
- Lastly, we will make Predictions based on our model.



Google I/O Extended

Thank You

Sandip Palit

Technology Consultant @ PwC

https://linkfree.io/SandipPalit