Leveraging AI and statistics to facilitate data-driven business decisions

Polly Wong

Driving change with quality statistics and data for everyone

Polly Wong

Chartered Statistician,
Royal Statistical Society, UK

Ex- Marketing Science Partner, Meta

- 15+ years of experience in AL/ML
- First ML model deployed in 2005 at IBM
- 4B+ USD business
  - Advertising Platform
  - eComm
  - Banking & Insurance
  - Autos
  - Hotel & Hospitality
  - Fintech
  - TV Shopping Channel
  - Publishing company
  - Energy
- Co Author for Meta Cross Border Business Guide 2022
- First Auction Expert representing APAC Regional at Meta

Statistics play a pivotal role on business decision making

OVERVIEW

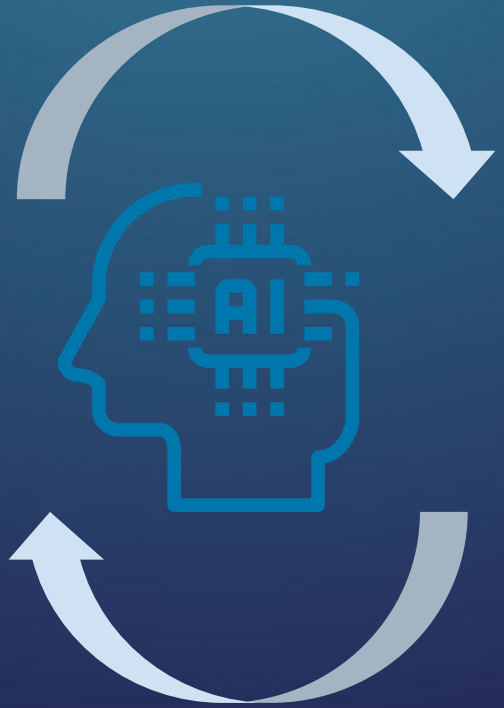
Statistics as foundation for AI

Regression & Classification

Unsupervised learning

Hypothesis testing

Probability Distributions



Predict fraud likelihood

Segment customers based on key factors

Validate assumptions

Identify outliers, calculate means

Data are everywhere

Online data

- SEO/SEM
- E-mail
- Transaction
- Social
- Web (Adobe)
- Mobile app

Structured data

Demographic e.g. gender, age
Transactional e.g. quantity, sales

Unstructured data

Image & video file
Chat history
Survey open-ended response data

Offline data

- Direct mail
- Broadcast
- Transaction
- Loyalty Card
- Call centers
- Customer / account (EDW)

Case Studies

Spotify

25%↑

Increase in retention rate

Collaborative FIltering matching profile with similar users and songs

amazon

20%↑

Improvement in Delivery time

Demand Forecasting Model & other techniques to predict customer demand and optimize inventory levels

slack

20%↓

Decrease in Churn rate

Logistic regression predicting accounts that were at risk of churning

Source: https://superagi.com/case-studies-in-ai-driven-customer-retention-success-stories-from-saas-ecommerce-and-subscription-services/#

How to create data-driven decisions

Quality data is crucial for successes

Business goals

Set the right KPI, goals and problem statements

Data Extraction

Extract data from different systems

Data Preparation

Run exploratory data analysis on duplicate, missing and accuracy

Model Build & Evaluation

Build with 70/30 training/ test data sets and evaluate with Accuracy/ Recall

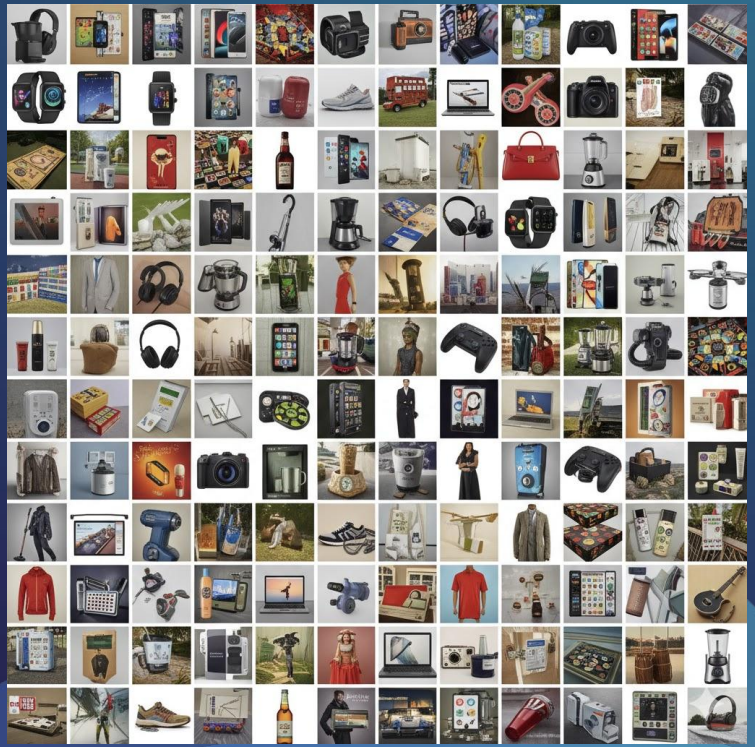
Data-driven action & Monitoring

Run hypothesis testing to validate the effectiveness of recommendations regularly

CASE 01

Optimizing advertising with LlaVA & Statistical Extreme Gradient Boosting model

Imagine eCommerce advertisers with 500K+ products



(illustration purpose only. AI-generated photo)

Context

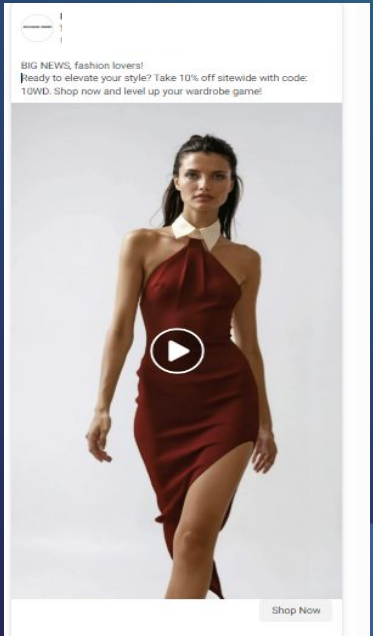
- 500K+ products that could be promoted online
- Most products with short product life-cycle (only popular for 2 months)
- Manual video is too costly

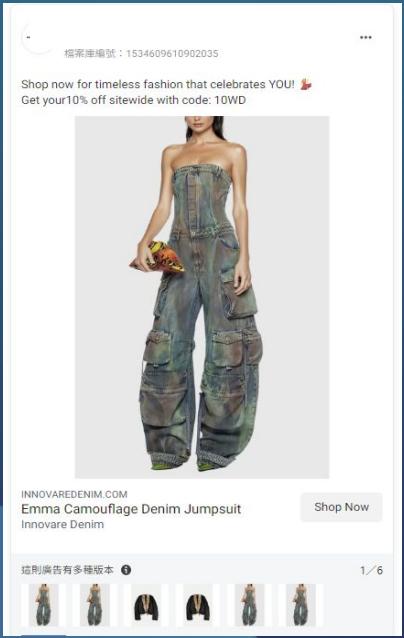
Problem statement

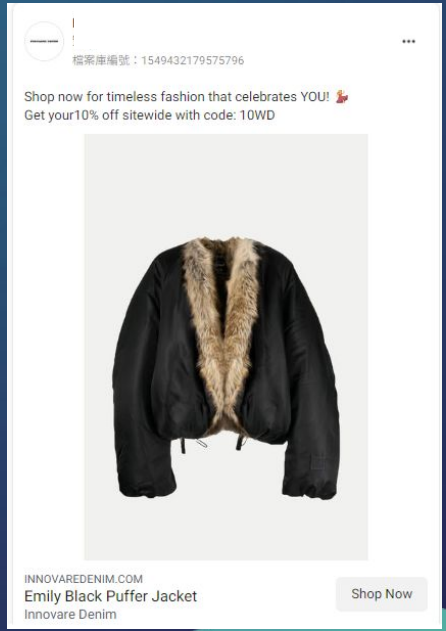
- How to identify which elements are performing and create effective ads for the right products at scale

Ad creatives look very different

Unstructured Image & video data
Huge data e.g. 30000+ products daily







Source: https://www.facebook.com/ads/library/ (illustration purpose only)

AI transform unstructured to structured features

- **Prompt AI LLaVA Large Language and Vision Assistant) to answer following questions**

Does Ad Have People (Yes/No)

Is Ad Static or not (Yes/No)

Does Ad have Music on or not (Yes/No)

Does Ad have Text overlay or not (Yes/No)

Does Ad have Background Color (Yes/No)

- **AI LLaVA produce 50+ ad features (structured)**
- **QA with sampling -70-80% accuracy**

(illustration purpose only)

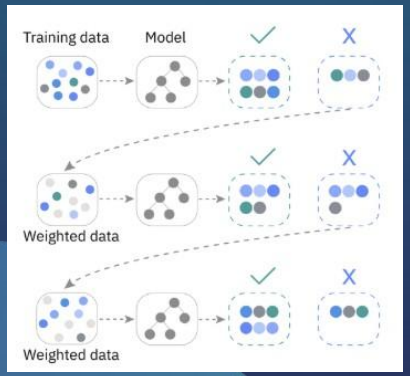
Find top contributing factors with statistical model

XGBoost (eXtreme Gradient Boosting)

supervised learning boosting algorithm that makes use of gradient descent

- **known for its speed, efficiency and ability to scale well with large datasets.**

Quality & structured datasets



Top important contributor

- **Music on**
- **Vertical Video**
- **Text Overlay**

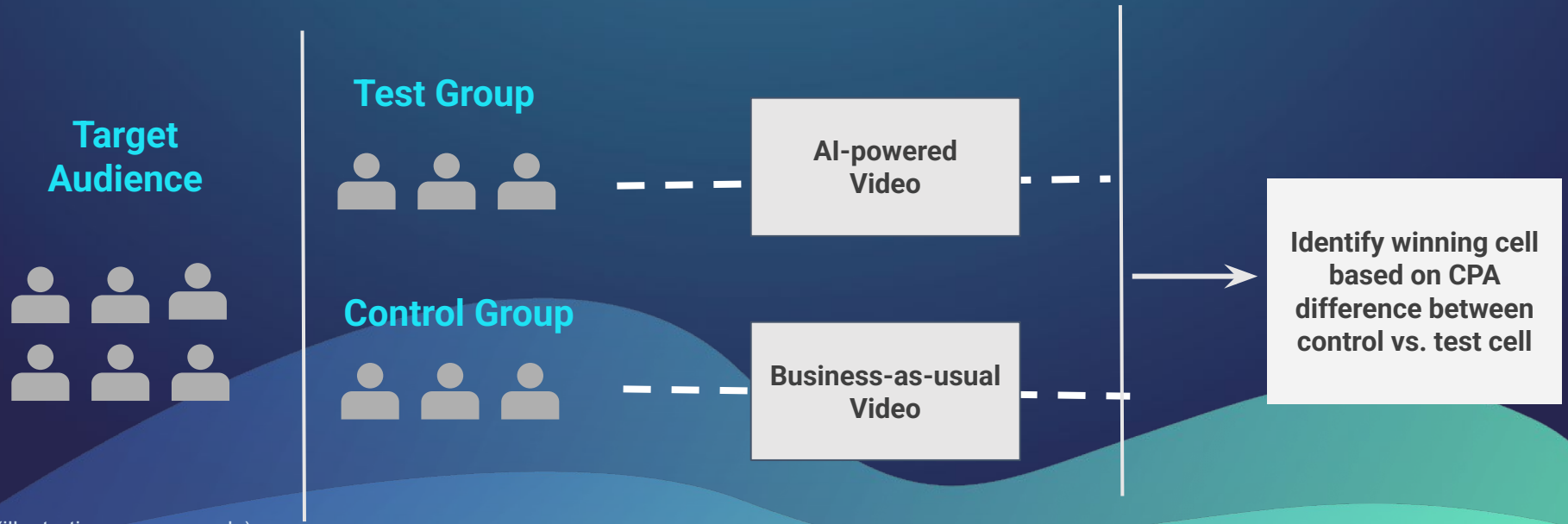
XGBoost (eXtreme Gradient Boosting)

(illustration purpose only)

Run hypothesis testing to validate new idea

Null Hypothesis: Business-as-usual video generate same Cost Per Purchase as AI-powered videos

T- Test showed test group generating statistical significant results at 95% C.I.



(illustration purpose only)

CASE 02

Increasing booking revenue with Customer Segmentation

Imagine hotel group with 50+ hotels in region

(illustration purpose only. AI-generated photo)

Background

- Hotel group have diverse properties (different star rating) with various focus/ facilities e.g. spa, gym, F&B, casino, resorts
- Regional office would like to standardize communication and one-solution-does not fit-all

Problem statement

- How many segments do we have in hotel groups
- How to design communication strategies for segments

Capture data along customer journey

Offline data

Online data

Create Single Customer View and data cleansing & transformation with new attributes

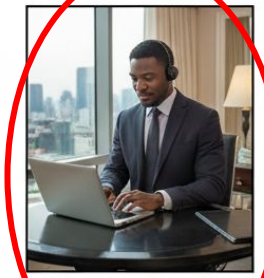
Inspire	Plan	Book	Experience	Return / Advocate
• Marketing campaigns • Mobile App notification	• Direct/3rd party booking (add to cart vs. free cancellation) • Booking w or w/o Cancellation pattern • Other hotel offering (weekly data)	• Guest • Booking (n day in adv/last minute) • Room/ Hotel type • Payment • Mobility (no. of unique hotels) • Weekday/Weekend • Purpose of travel • Chat history	• Usage: SPA, Pool, Casino F&B • Catering (In room service, restaurant) • Interest class e.g. Running, Yoga, Culinary • Lifestyle (sustainable travel)	• Hotel/chain location level • Frequency • Spending • Recency • Regularity • Loyalty Tier

Booking data is not perfect - booking is made under different persons for same traveller/family

Aggregate data from 3rd party booking site (weekly) is not mapped to individual booking (daily)

Spa Lover

Gym/Pool Goer



Business Traveller

Couples

Couples with Kids

Family with Elderly

Budget Travelers

F&B Lover

Celebrity

Yoga Doers

Last-Minute Booker

Travel Agency

K-mean clustering leads to 12 segments

- Business vs Leisure is important differentiator on needs
- Targeted offers for each segments
- Travel agency: special segment with very unique needs)

Business traveler offers (Focus: Productivity, convenience, and efficiency)

- Welcome drinks
- Free access to executive lounges with meeting rooms and printing

Run hypothesis testing to validate new idea

Null Hypothesis: General offers generate same conversion rate and click through rate as segmented offers

T- Test showed statistical significant results, 3x higher CTR comparing segmented-offers vs. general offers

Email Conversion rate = Booking/ No. of delivered emails. Email Click through rate = Clicks/ No. of delivered emails

Create heat map to identify opportunities

Split by Hotel	Hotel A	Rev %	Avg No. of days between booking	Email Contactable %
Business traveller	30%	50%	14	95%
Celebrity	1%	5%	60	62%
Spa lover	3%	1%	320	56%
Yoga folks	0%	0%	0	33%
F&B Foodie	8%	3%	50	55%
Gym/Pool Goer	1%	1%	0	76%
Couple with kids	3%	2%	114	23%
Couple with elderly	9%	8%	80	11%
Couple	18%	14%	45	45%
Budget travellers	3%	2%	183	89%
Last minute booker	8%	3%	28	55%
Travel agency	16%	11%	5	99%

- Find opportunity area with monthly Statistics to facilitate rebooking with target offers (email) at right segment
  - Business traveller - core revenue stream
  - Gym/Pool goer - deep dive
- Projected improvement on rebooking by 8%-16% (6 month evaluation window)

(illustration purpose only)

Wrap up

- ☑ AI and Statistics is a game changer for smarter decisions
- ☑ Always check data quality and model fairness/explainability
- ☑ Quality data remains critical anytime anywhere.
  Data-driven business decising only work with quality statistics

THANK YOU !