

# Extreme Classification

COV 878: Special Topics in  
Machine Learning

Manik Varma

Microsoft Research & IIT Delhi



# Binary Classification

---

- Answer yes/no questions involving uncertainty



Is this George Washington or not?

# Multi-class Classification

---

- Answer multiple choice questions



Which US President is present in this image?



# Multi-label Classification

---

- Pick multiple answers in a multiple choice question



Which US Presidents are present in this image?

# Traditional Classification

- Classification with a small number of choices



# Extreme Classification

- Classification with millions of labels

Ad



Do you know how much you could save on car insurance by switching to GEICO?

Get an online quote:  [GO](#)

Get a quote by phone: [1-800-841-1588](tel:1-800-841-1588)

Find a local agent:  [GO](#)

Need other insurance?

ATV	Homeowners	Overseas
Boat	ID Theft Protection	Renters
Commercial Auto	Life	RV
Condo/Ca-op	Mobile Home	Umbrella
Flood	Motorcycle	

Already a policyholder? [Log in.](#)

**GEICO**

© 1996-2012 GEICO

Predictions

geico auto insurance

geico car insurance

geico insurance

www geico com

care geicos

geico com

need cheap auto insurance

wisconsin cheap car insurance quotes

cheap auto insurance florida

all state car insurance coupon code

# Extreme Classification

---



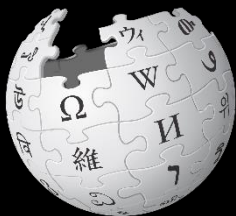
- Publications at AAAI, AISTATS, ECCV, ICML, KDD, NIPS, SIGIR, WSDM, WWW, *etc.*



- 8 popular workshops organized in 5 years at Dagstuhl, ECML, ICML, NIPS, WWW, *etc.*



- Code, datasets & benchmarks released on The Extreme Classification Repository



- Wikipedia results have improved from 20% in 2013 to 65% in 2017



# Applications

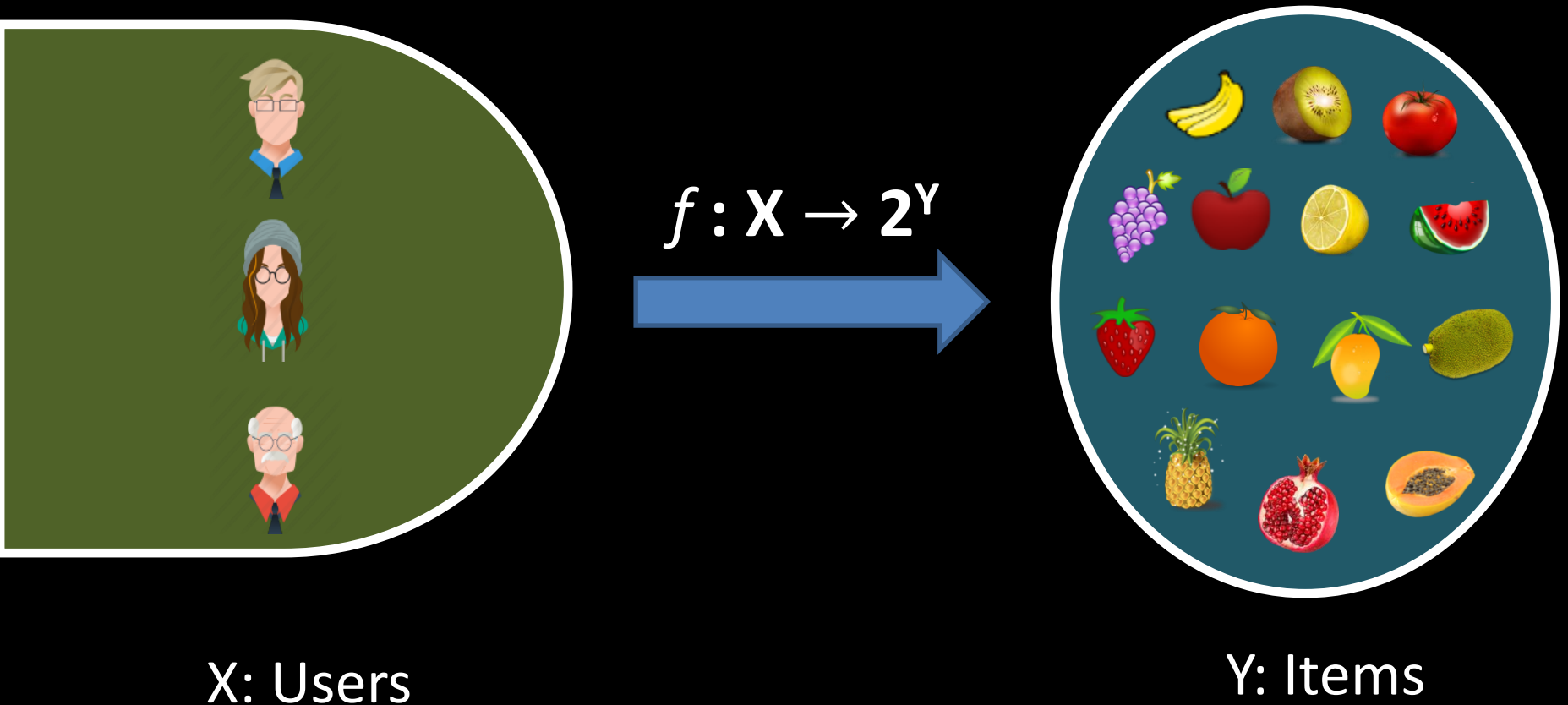
- Information Retrieval
  - Ranking for web search & advertising
- Recommender Systems
  - Item to item recommendation
- Natural Language Processing
  - Language modelling
  - Document tagging
- Computer Vision
  - Person recognition
  - Learning universal feature representations
- Bioinformatics
  - Gene function prediction





# Extreme Multi-Label Classification

- Problem formulation

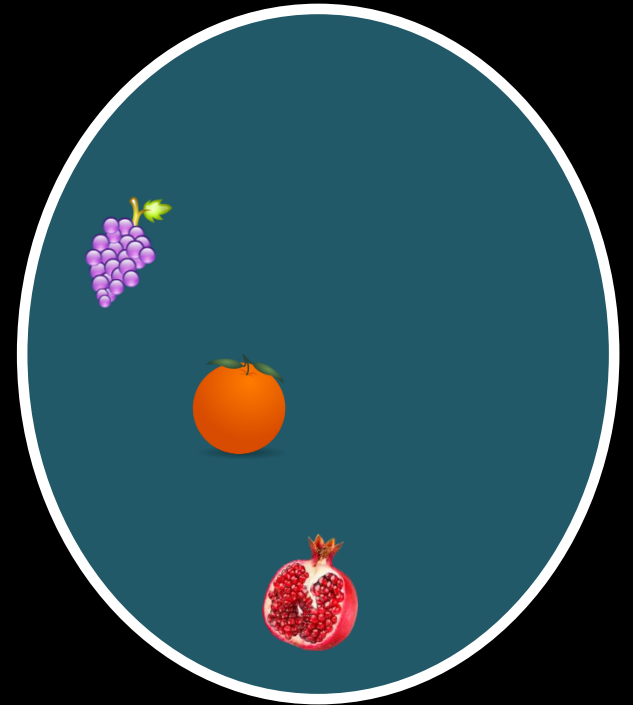


# Extreme Multi-Label Learning


---

- Problem formulation

$f$  (  )



# Bing Ads – Tesco's Distilled Water

**amazon.co.uk** [Try Prime](#) All ▾ distilled water 

Shop by **Department ▾** Your Amazon.co.uk Today's Deals Christmas Store Hello. Sign in **Your Account ▾** Try **Prime ▾** Your **Lists ▾** **0** **Basket**

[Amazon.co.uk](#) [Today's Deals](#) [Warehouse Deals](#) [Outlet](#) [Subscribe & Save](#) [Vouchers](#) [Amazon Family](#) [Amazon Prime](#) [Amazon Pantry](#) [Amazon Video](#)

1-16 of 3,179 results for "distilled water" Sort by **Relevance** ▾



**Show results for**

**Automotive**

- Vehicle Oils & Fluids
- Winter Goods for Cars
- Car Lighting Assemblies & Components

**Health & Personal Care**

- Wellness
- Medication & Remedies
- Vitamins, Minerals & Supplements
- Alternative Medicine
- Colloidal Silver
- [See more](#)

	<p><b>Sponsored</b></p> <p><b>Distilled Water - 5 Litre (5L) - Pure Steam Distilled - 0ppm</b> by Distilled Water UK</p> <p><b>£24.95</b> (£4.99/l) <span style="float: right;">30</span></p>
	<p><b>Adamol 1896 51635139 Distilled Water 5 L</b> by ADAMOL 1896</p> <p><b>£7.20</b> <span style="float: right;">16</span></p>

**Bidded Query: distilled water 5 litres**

# Predictions: Bing Ads vs Extreme Classification

---

## Bing Ads

water 5

## Extreme Classification

distilled water tesco

where buy distilled water

distilled water

buy distilled water

distilled water amazon

distilled water vs purified water

distilled water uk

distilled water delivery

where can I buy distilled water

distilled water uk supermarket



# Traditional Approach

---

- Reduction to binary classification

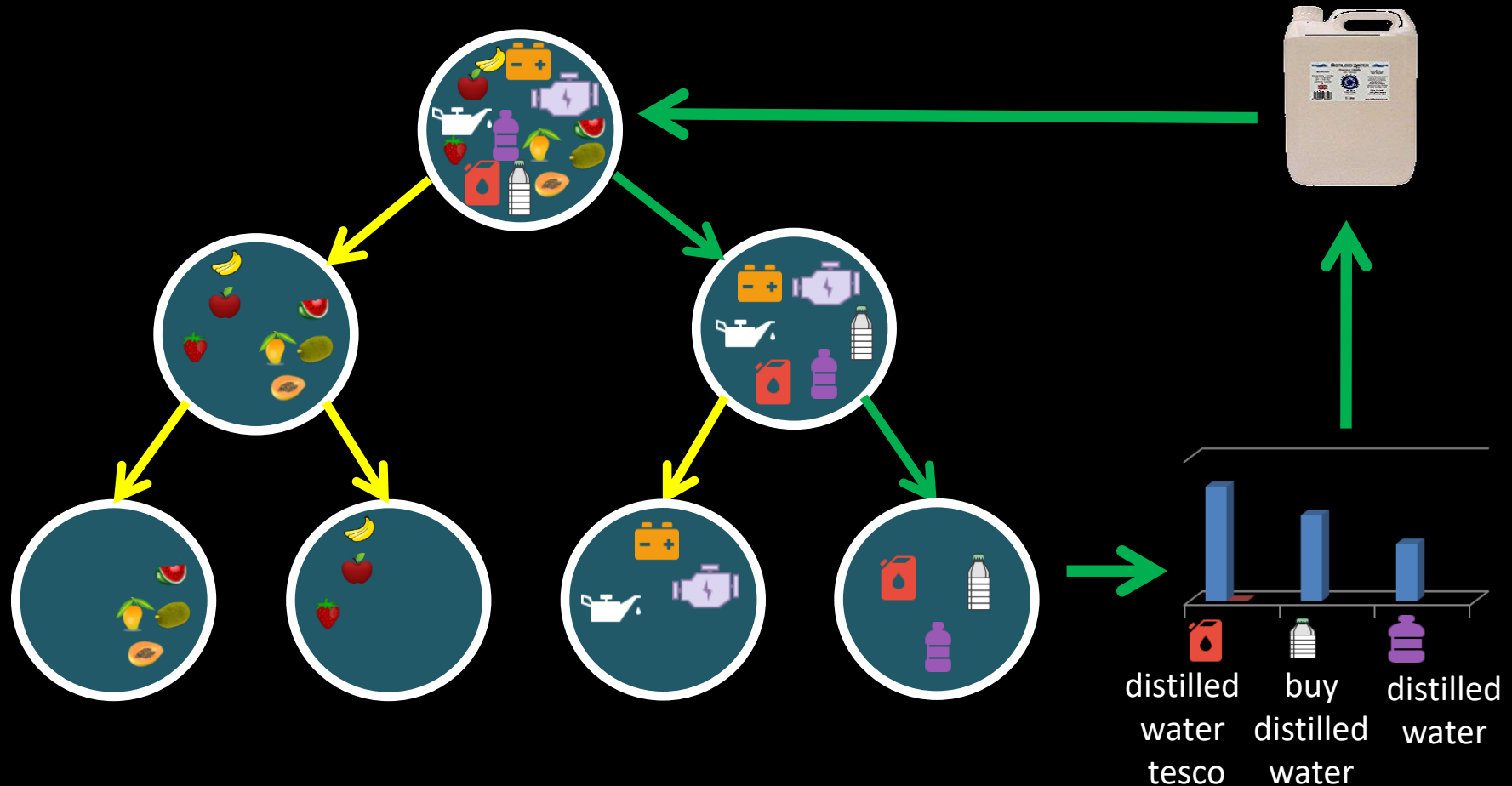
$h : (\text{Ad}, \text{Phrase}) \rightarrow \{\text{✗}, \text{✓}\}$

$h(\text{🛒}, \text{buy distilled water}) \rightarrow \text{✗}$

$h(\text{🛒}, \text{water 5}) \rightarrow \text{✓}$

# Extreme Classification Approach

- Efficient & accurate prediction via a learnt hierarchy



# Extreme Classification for Bing Ads

## UK Dynamic Search Ads

The screenshot shows a product page for 'Impey Coving Ext Profile Vinyl' on the Topps Tiles website. The page features a large image of the vinyl product, a price tag of £54.65, and a quantity selector set to 0. Below the product image, there is a section for 'Additional Information' which includes a description: 'The Impey Coving Ext Profile is a single piece of anodised aluminium that accepts the cover frame on a vinyl floor.' It also lists 'Availability' as 'Specialty Orders For You' and 'Delivery options' as 'Home or Store'.

## French Text Ads

The screenshot shows a French insurance website with a yellow highlight on the bottom text. The website lists various insurance services such as 'Assurance Vie CIBC', 'Assurance Auto', and 'Assurance Moto'. The highlighted text at the bottom reads: 'Bided Keywords: la vie assurance, assurance auto, assurance moto'.

## German Product Ads

The screenshot shows a German product advertisement for a Nikon Coolpix B500 camera. The ad features a high-quality image of the red camera. Below the image, the text reads: 'Nikon Coolpix B500 Pflaume' and '259,00 €'. At the bottom, it mentions 'Kostenloser Versand' (Free shipping) and 'keine Versandinfo' (no shipping info).

# Product Recommendation on Amazon

Books › History › Americas

## Rare Wildflowers of Kentucky Hardcover – August 1, 2008

by Thomas G. Barnes (Author), Deborah White (Author), Marc Evans (Author)

★★★★★ 1 customer review

Look inside ↴



▶ See all 3 formats and editions

Kindle  
\$34.36

Hardcover  
**\$39.95**

Read with Our **Free App**

12 Used from \$21.97

20 New from \$19.85

1 Collectible from \$17.95

Kentucky's rich soil and wonderfully diverse landscapes have for centuries made the state a welcoming habitat for a dazzling variety of wildflowers. From the delicate Kentucky glade cress to the fiery royal catchfly, flowers, grasses, and sedges nestle on rocky hillsides, bask in open woodlands and prairies, and even manage to thrive in busy rights of way and roadsides. As demand for natural resources and land for

▶ Read more



See all 2 images



# Predictions: Amazon vs Extreme Classification

## Amazon

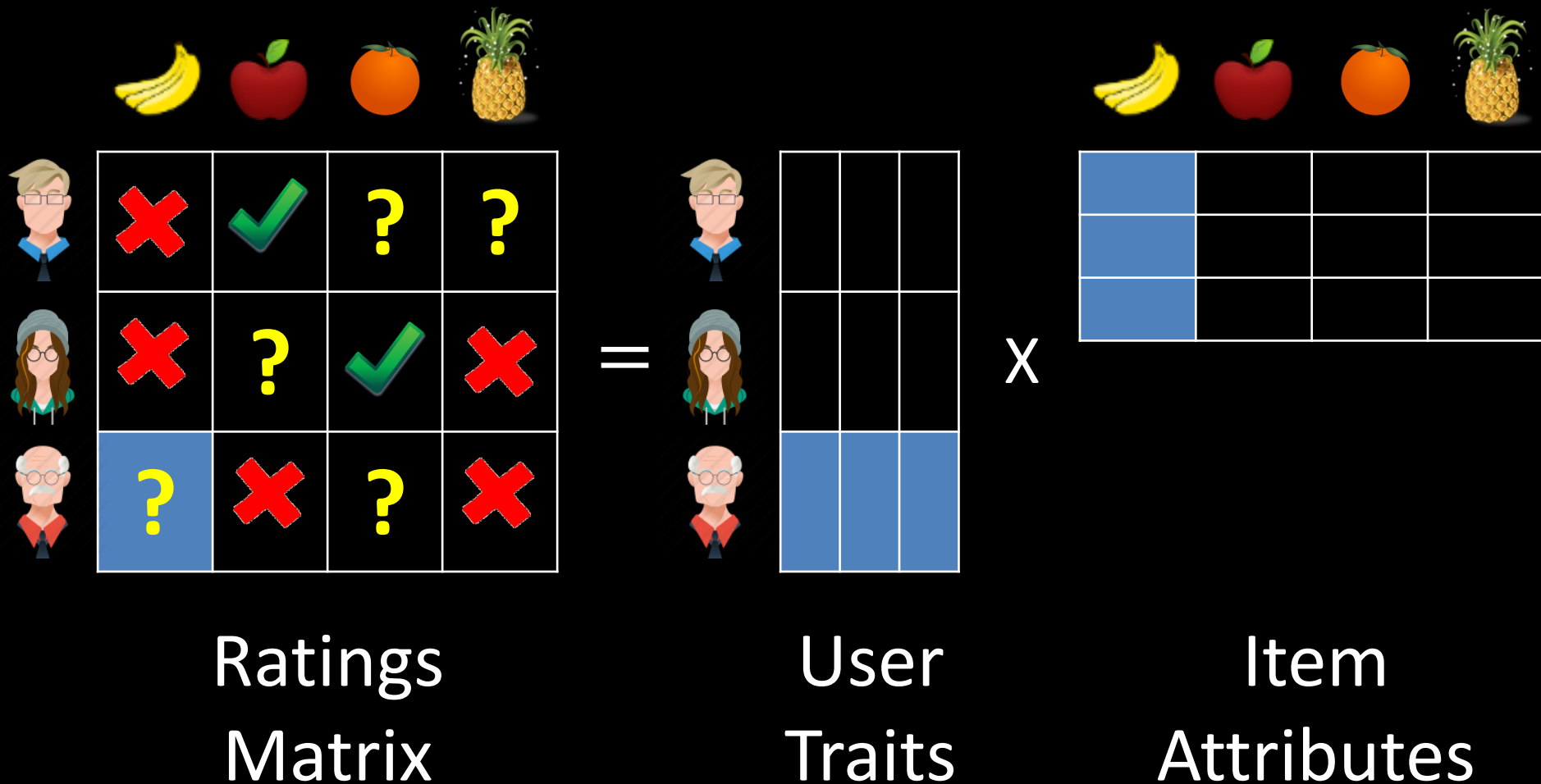
<b>Kentucky's Last Great Places</b>
<b>Trees and Shrubs of Kentucky</b>
<b>Wildflowers and Ferns of Kentucky</b>

## Extreme Classification

<b>Trees and Shrubs of Kentucky</b>
<b>Kentucky's Last Great Places</b>
<b>Kentucky Trees &amp; Wildflowers: A Folding Pocket Guide to Familiar Species</b>
<b>Birds of Kentucky Field Guide</b>
<b>Woody Plants of Kentucky &amp; Tennessee: The Complete Winter Guide to Their Identification &amp; Use</b>
<b>Native Trees of Kentucky: A Handbook</b>
<b>Kentucky Wildlife: A Folding Pocket Guide to Familiar Species</b>
<b>Wildflowers and Ferns of Kentucky</b>
<b>Kentucky's Natural Heritage: An Illustrated Guide to Biodiversity</b>
<b>Kentucky Birds: A Folding Pocket Guide to Familiar Species</b>

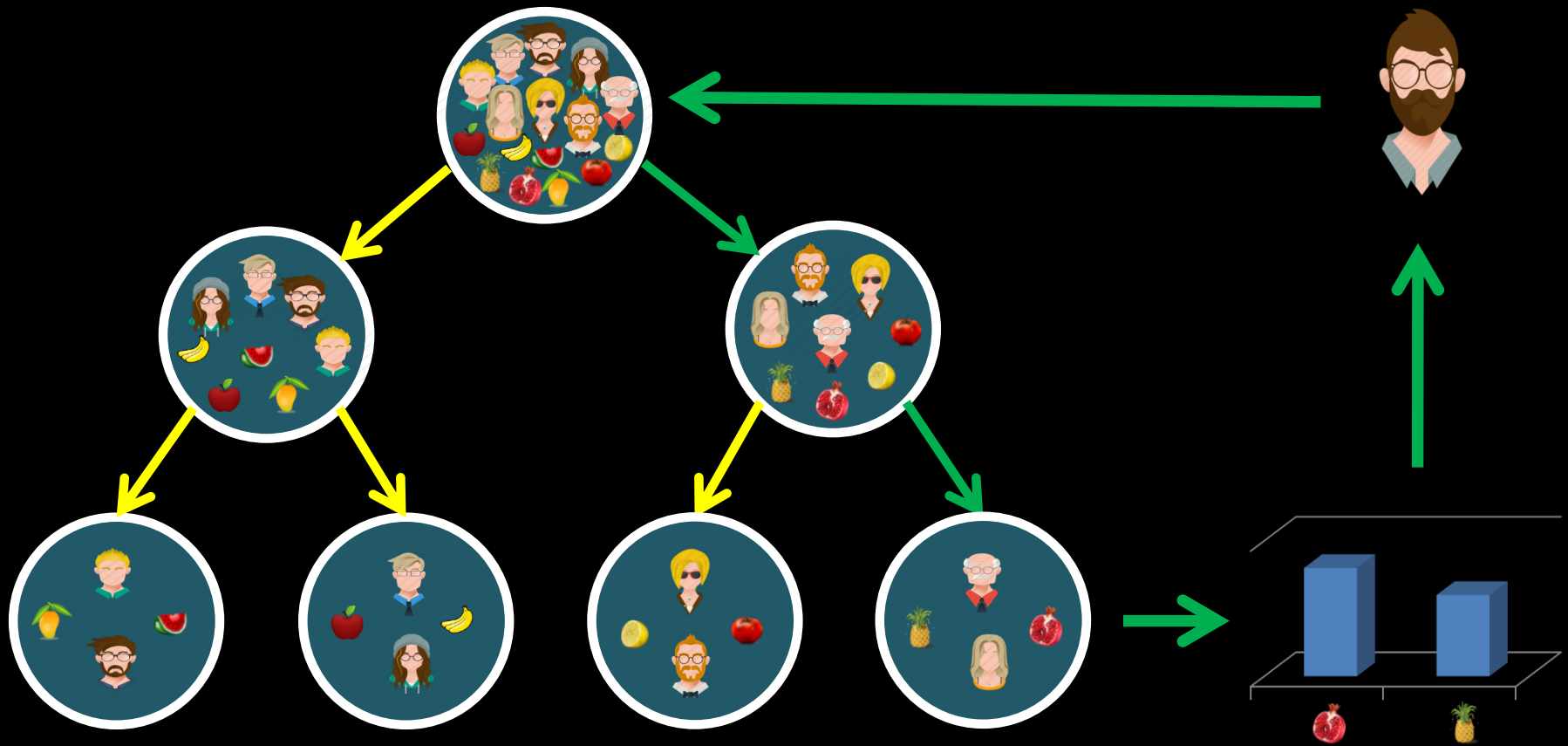
# Traditional Approach

- Collaborative filtering & matrix factorization



# Extreme Classification Approach

- Recommendation based on user and item features



SwiftXML [Prabhu, Kag, Gopinath, Harsola, Agrawal, Varma WSDM 18]

# Bing RS – “cam procedure shoulder”

cam procedure shoulder



Web

Images

Videos

Maps

News

My saves

37,200,000 Results

Any time ▾

## CAM Procedure | Shoulder Joint Preservation | Osteoarthritis

[drmillett.com/...restoration-procedures-for-the-shoulder-cam-procedure ▾](#)

Dr. Peter J. Millett developed the CAM (Comprehensive Arthroscopic Management) procedure for osteoarthritis of the shoulder.

Related searches

cam newton shoulder surgery

## Comprehensive Arthroscopic Management - CAM Procedure ...

[www.prweb.com/releases/delayshoulderreplacement/camprocedurefor... ▾](#)

Vail, Colorado (PRWEB) September 20, 2013 -- Newly published research, shows that the innovative CAM procedure is a viable surgical option to delay shoulder

## [VIDEO] CAM Procedure | Shoulder Osteoarthritis Surgery | Vail and ...



[www.youtube.com/watch?v=yqi-2dNJ\\_0A ▾](#)

By shoulderspecialists · 5.4K views

Nov 30, 2010 · [http://drmillett.com](#) Dr. Peter J. Millett is an orthopedic shoulder surgeon with the Steadman Clinic in Vail, CO and is the pioneer of the CAM procedure.



# Predictions: Bing vs Extreme Classification

---

## Bing

cam newton shoulder surgery

## Extreme Classification

types of shoulder surgical procedures

shoulder replacement lawsuits

how long off work for shoulder surgery

common shoulder surgeries

stem cell therapy for rotator cuff tear

what to wear after shoulder surgery

cost of arthroscopic shoulder surgery

arthroscopic shoulder surgery

# Tagging Wikipedia Articles



Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article **Talk**

[Read](#) [Edit](#) [View history](#)

Search Wikipedia



**WIKIPEDIA**  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)  
[Wikipedia store](#)

[Interaction](#)

[Help](#)  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)  
[Contact page](#)

[Tools](#)

[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Permanent link](#)

## *Divine Comedy*

From Wikipedia, the free encyclopedia

*"The Divine Comedy"* redirects here. For other uses, see *The Divine Comedy (disambiguation)*.

*"La commedia"* redirects here. For other uses, see *Commedia (disambiguation)*.

The ***Divine Comedy*** (Italian: *Divina Commedia* [diˈviːna komˈmɛːdja]) is a long narrative poem by Dante Alighieri, begun c. 1308 and completed in 1320, a year before his death in 1321. It is widely considered the preeminent work in Italian literature,<sup>[1]</sup> and is seen as one of the greatest works of world literature.<sup>[2]</sup> The poem's imaginative vision of the afterlife is representative of the medieval world-view as it had developed in the Western Church by the 14th century. It helped establish the Tuscan language, in which it is written, as the standardized Italian language.<sup>[3]</sup> It is divided into three parts: *Inferno*, *Purgatorio*, and *Paradiso*.

The narrative describes Dante's travels through Hell, Purgatory, and Paradise or Heaven;<sup>[4]</sup> while allegorically the poem represents the soul's journey towards God.<sup>[5]</sup> Dante draws on medieval Christian theology and philosophy, especially Thomistic philosophy and the *Summa Theologica* of Thomas Aquinas.<sup>[6]</sup> Consequently, the *Divine Comedy* has been called "the *Summa* in verse".<sup>[7]</sup>



Dante shown holding a copy of the *Divine Comedy*, next to the entrance to Hell, the seven terraces of Mount Purgatory and the city of Florence, with the spheres of Heaven above, in Michelino's fresco

# Predictions: Wiki vs Extreme Classification

---

## Wikipedia

<b>Works by Dante Alighieri</b>
<b>Divine Comedy</b>
<b>1321 books</b>
<b>1300 in Italy</b>
<b>Visionary poems</b>
<b>Epic poems in Italian</b>
<b>14th-century Christian texts</b>
<b>14th-century books</b>
<b>Virgil</b>
<b>Afterlife</b>

## Extreme Classification

<b>Works by Dante Alighieri</b>
<b>Divine Comedy</b>
<b>1321 books</b>
<b>1300 in Italy</b>
<b>Visionary poems</b>
<b>Epic poems in Italian</b>
<b>14th-century Christian texts</b>
<b>14th-century books</b>
<b>Virgil</b>
<b>Dante Alighieri</b>



# Recognizing People on Facebook



Choices: Bradley Cooper, Ellen DeGeneres, Meryl Streep, Jennifer Lawrence, Channing Tatum, Julia Roberts, Kevin Spacey, Brad Pitt, Angelina Jolie, Lupita Nyong'o, Peter Nyong'o



# Language Modelling

---

Brevity is the soul of ...



Wit



Twit



Lingerie

# Conclusions

- Extreme classification
  - Tackle applications with millions of choices
  - A new paradigm for ranking & recommendation
- Algorithms & papers
  - MLRF [WWW 2013], FastXML [KDD 2014]
  - SLEEC [NIPS 2015], PfastreXML[KDD 2016]
  - SwiftXML [WSDM 2018], Parabel [WWW 2018]
- The Extreme Classification Repository
  - Code & datasets
  - Benchmark results
  - Papers





# Research Questions

- Applications
- Obtaining good quality training data
- Log time and space training and prediction
- Obtaining discriminative features at scale
- Extreme loss functions
- Performance evaluation
- Dealing with tail labels and label correlations
- Dealing with missing and noisy labels
- Explore/exploit for tail labels
- Statistical guarantees
- Fine-grained classification