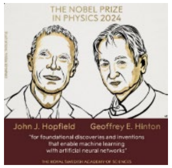
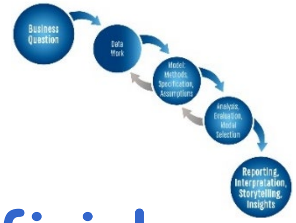


Leadership Initiatives

AI & Emerging Technology Internship



Predictive Analytics, Machine Learning & Artificial Intelligence for Business: Key Concepts and Trends

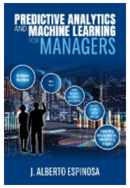
Prof. J. Alberto Espinosa, PhD

Dept. of Information Technology & Analytics

alberto@american.edu

November 6, 2024

Download slides:
https://tinyurl.com/PUCPEspinosa



Ch1 R: GitHub

Book ChatBot
https://poe.com/ITEC-621

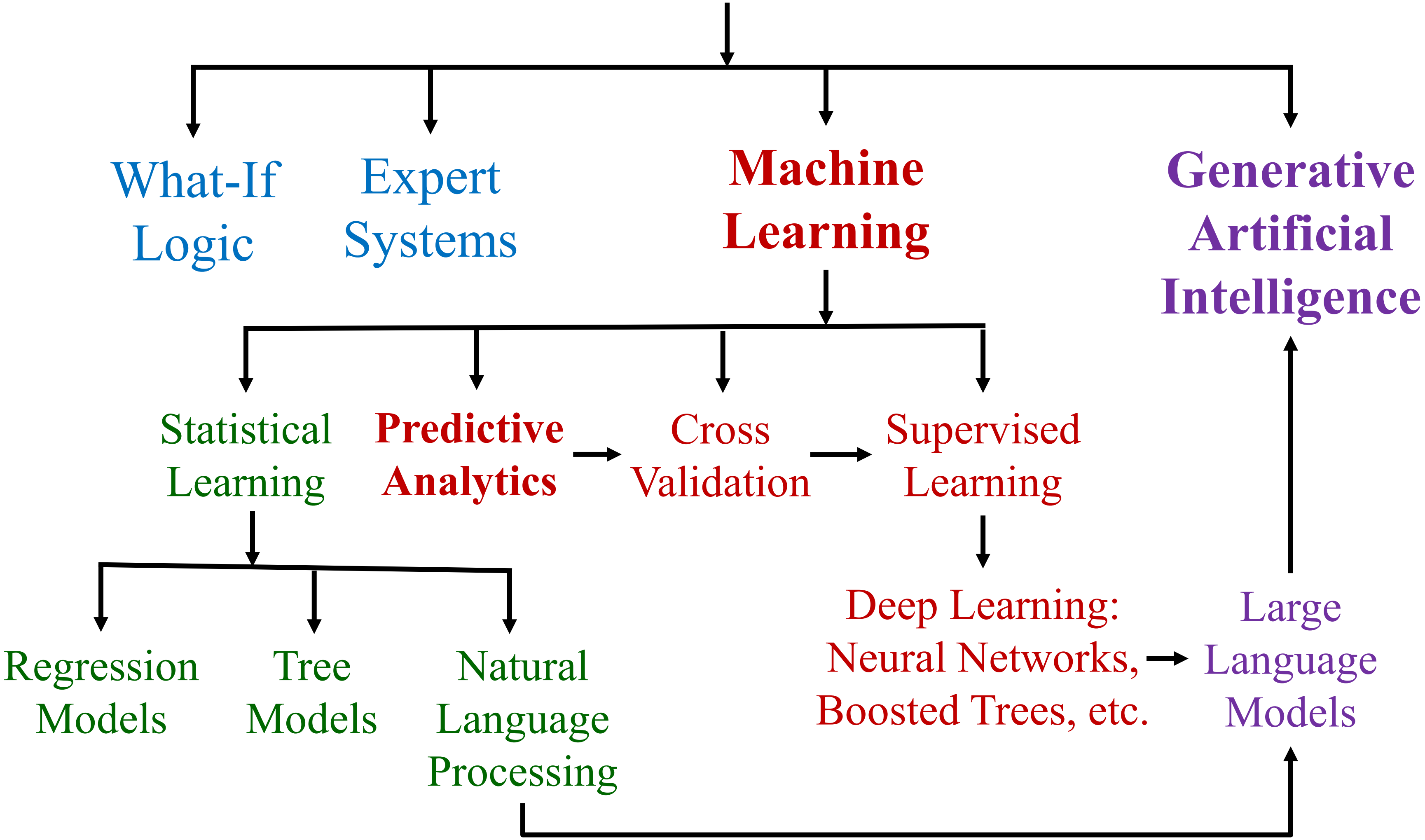
KOGOD SCHOOL of BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

J. Alberto Espinosa, Ph.D.

Professor of Information Technology & Analytics
Kogod School of Business, American University, Washington, D.C.

Prof. Espinosa holds a Ph.D. and Master of Science in Information Systems from the Tepper School of Business at Carnegie Mellon University, a Masters in Business Administration from Texas Tech University; and a Mechanical Engineering degree from Pontificia Universidad Catolica del Peru. He is the architect of Kogod's MS Analytics degree and the undergraduate specialization in Business Analytics. He is also the curriculum architect for the IT&A programs and specializations. He has over 70 published journal articles, conference proceedings, books and book chapters. He has co-authored two research books, one on work coordination across time zones and a 2-volume book on big data and analytics for service delivery. He has also written a book titled "Predictive Analytics and Machine Learning for Managers" with practical applications in R. His research focuses on coordination and performance in technical projects across global boundaries, particularly distance and time zones. More recently he has developed social network and graph analytics methods to represent team knowledge quantitatively and visually. His work has been published in leading scholarly journals, including Management Science, Organization Science, Information Systems Research, the Journal of Management Information Systems, IEEE Transactions on Software Engineering, IEEE Transactions on Engineering Management, Communications of the ACM, Human Factors, Information, Technology and People, and Software Process: Improvement and Practice. His work has also been featured in leading academic conference proceedings. He teaches predictive analytics with machine learning, social and organizational network analytics, R programming, information technology foundations, business process analysis, and programming for business applications. He also has several years of working experience, first as a mechanical design engineer and later as a senior manager, VP, and CFO with international organizations directly supporting, supervising, and formulating policy for finance and global IT, data management and analysis, where he designed and developed a number of software applications to support geographically distributed work in accounting, finance and program monitoring and evaluation.

Predictive Analytics, Machine Learning & Artificial Intelligence for Business



What-If Logic

Expert Systems

Machine Learning

Generative Artificial Intelligence

Statistical Learning

Predictive Analytics

Cross Validation

Supervised Learning

Regression Models

Tree Models

Natural Language Processing

Deep Learning: Neural Networks, Boosted Trees, etc.

Large Language Models

Predictive Analytics

THE WALL STREET JOURNAL.
WSJ.com

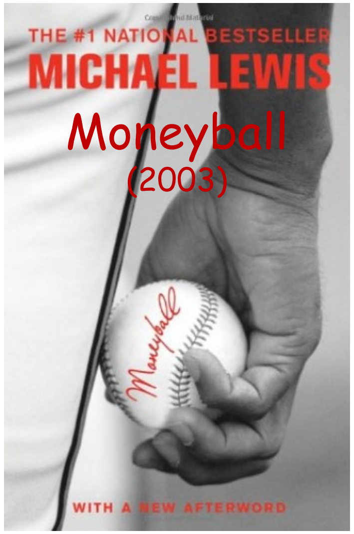
September 11, 2015, 9:28 AM ET

Facebook 'Likes' Mean a Computer Knows You Better Than Your Mother

ByGeorgia Wells

A computer needs the data from only 10 Facebook "likes" to beat the accuracy of a person's coworker at judging his or her personality traits, such as extraversion, conscientiousness and neuroticism. The computer needs 70 "likes" to be more accurate than a person's friends. With 250 Facebook data points, the computer can beat someone's spouse.

Business Analytics



Moneyball
(2003)

"It is the scientific process of transforming data into insight for making better decisions" (INFORMS)

Lack analytical talent – by 2018, the US faces a shortage of **140,000** to **190,000** people with deep analytical skills and **1.5 million** managers with analytical skills to make decisions with data (McKinsey Global Institute, 2011)

The Age of Analytics
McKinsey Global Institute
2016

➡ **2M-4M**

Projected US demand for "business translators" in the 2020 decade

High demand for **business translators** → people who serve as the link between data science talent and business applications and managers, who are business-savvy and can answer business questions.

The Analytics Process

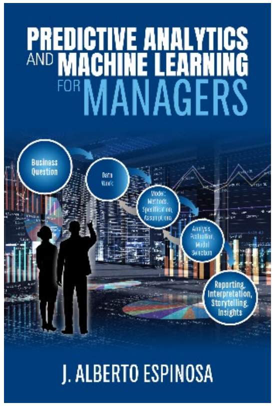
Business Question

Data Work

Model: Methods, Specification, Assumptions

Analysis, Evaluation, Model Selection

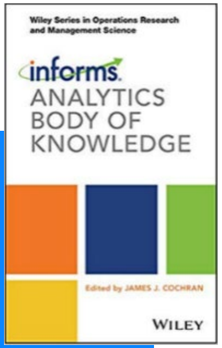
Reporting, Interpretation, Storytelling, Insights



PREDICTIVE ANALYTICS AND MACHINE LEARNING FOR MANAGERS

J. ALBERTO ESPINOSA

Analytics Body of Knowledge



Wiley Series in Operations Research and Management Science

informs.

ANALYTICS BODY OF KNOWLEDGE

Edited by JAMES J. COCHRAN

WILEY

Business Domains

- Business
- Healthcare
- Marketing
- Social Networks
- Accounting, Fraud Detection & Forensics
- Financial Analysis
- Data Governance
- Sports
- Cyber Security

Analytics Core

- Descriptive Analytics
- Prescriptive Analytics
- Machine Learning
- Data Mining
- Artificial Intelligence
- Predictive Analytics
- Text Mining
- Data Visualization
- Business Intelligence

Foundations (IT)

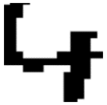
- Information Technology
- Database & Big Data
- Software & Tools
- Data Warehousing
- Math & Statistics
- Programming R, Python

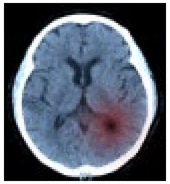
Machine Learning

Machine Learning

Is method of data analysis that uses algorithms that **learn from data** iteratively allowing computers to find hidden insights **without being explicitly programmed** (SAS: http://www.sas.com/en_us/insights/analytics/machine-learning.html)

- **Unsupervised** learning → data exploration **without** specific **goals** (closely associated with descriptive analytics and data mining, e.g., clustering, correlation analysis, histograms)
- **Supervised** learning → data analysis **with** specific **goals** in mind (closely associated with analytics, e.g., regression) → not much different than **predictive modeling**





$600

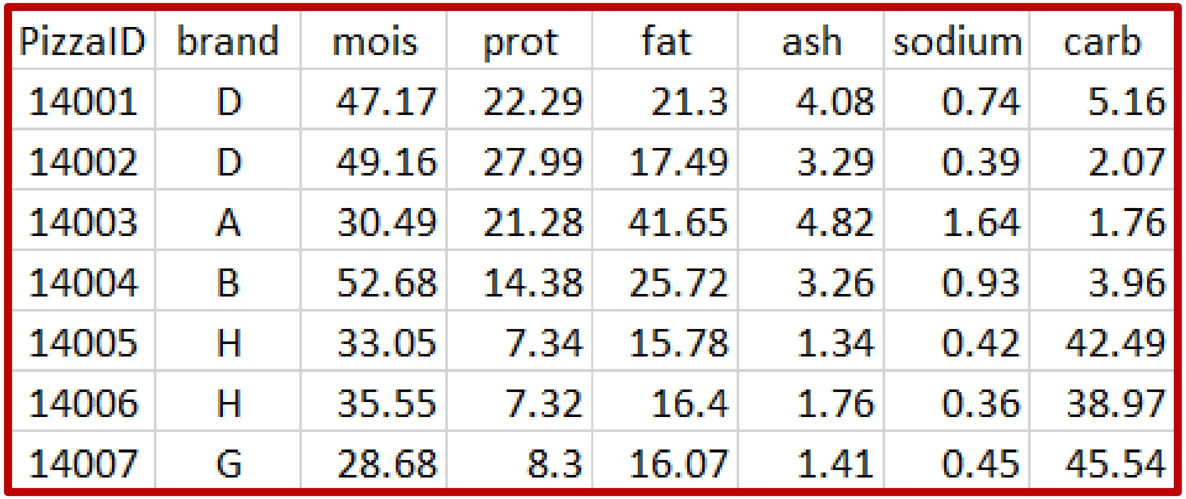
$200 OFF
Miele CM7750 Fully Automatic Espresso Machine

$5,300

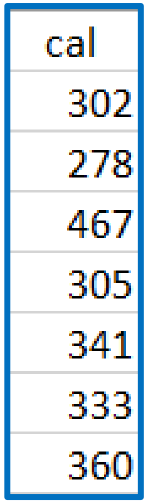
Predictive Model

Without Machine Learning & Cross Validation

Predictors → ***Outcome***



PizzaID	brand	mois	prot	fat	ash	sodium	carb
14001	D	47.17	22.29	21.3	4.08	0.74	5.16
14002	D	49.16	27.99	17.49	3.29	0.39	2.07
14003	A	30.49	21.28	41.65	4.82	1.64	1.76
14004	B	52.68	14.38	25.72	3.26	0.93	3.96
14005	H	33.05	7.34	15.78	1.34	0.42	42.49
14006	H	35.55	7.32	16.4	1.76	0.36	38.97
14007	G	28.68	8.3	16.07	1.41	0.45	45.54



cal
302
278
467
305
341
333
360

Machine Learning

Cross-Validation → Split, Train, Test, Re-Sample

Training Subset: Train Model →

PizzaID	brand	mois	prot	fat	ash	sodium	carb
14001	D	47.17	22.29	21.3	4.08	0.74	5.16
14002	D	49.16	27.99	17.49	3.29	0.39	2.07
14003	A	30.49	21.28	41.65	4.82	1.64	1.76
14004	B	52.68	14.38	25.72	3.26	0.93	3.96

Outcome

cal
302
278
467
305

Testing Subset: Test Model's Accuracy →

14005	H	33.05	7.34	15.78	1.34	0.42	42.49
14006	H	35.55	7.32	16.4	1.76	0.36	38.97
14007	G	28.68	8.3	16.07	1.41	0.45	45.54

Outcome

Actual	***Predicted***	*Error*
341	290	51
333	362	-29
360	371	-11

Mean Squared Error: 1187.7
RMSE = 34.5

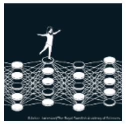
Selected Model

New Data

D	49.16	27.99	17.49	3.29	0.39	2.07

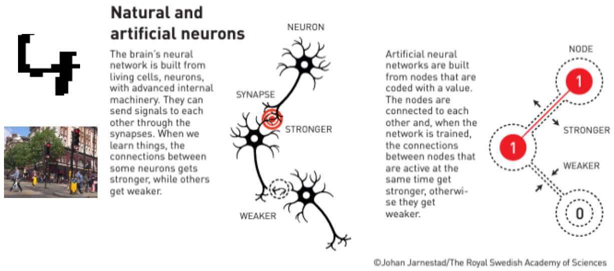
Predict:
- *Values*
- *Classifications*
- *Words*
- *Images, etc.*

Uses: test and compare models, tune model parameters, train deep and supervise learning models, etc.



Machine Learning Models

- **Simple learning** → simple linear relationships & 1 or 2 layers
- **Deep learning** → linear or non-linear relationships & multiple layers

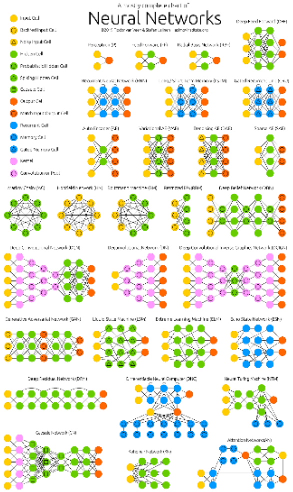


Natural and artificial neurons

The brain's neural network is built from living cells, neurons, with advanced internal machinery. They can send signals to each other through the synapses. When we learn things, the connections between some neurons gets stronger, while others get weaker.

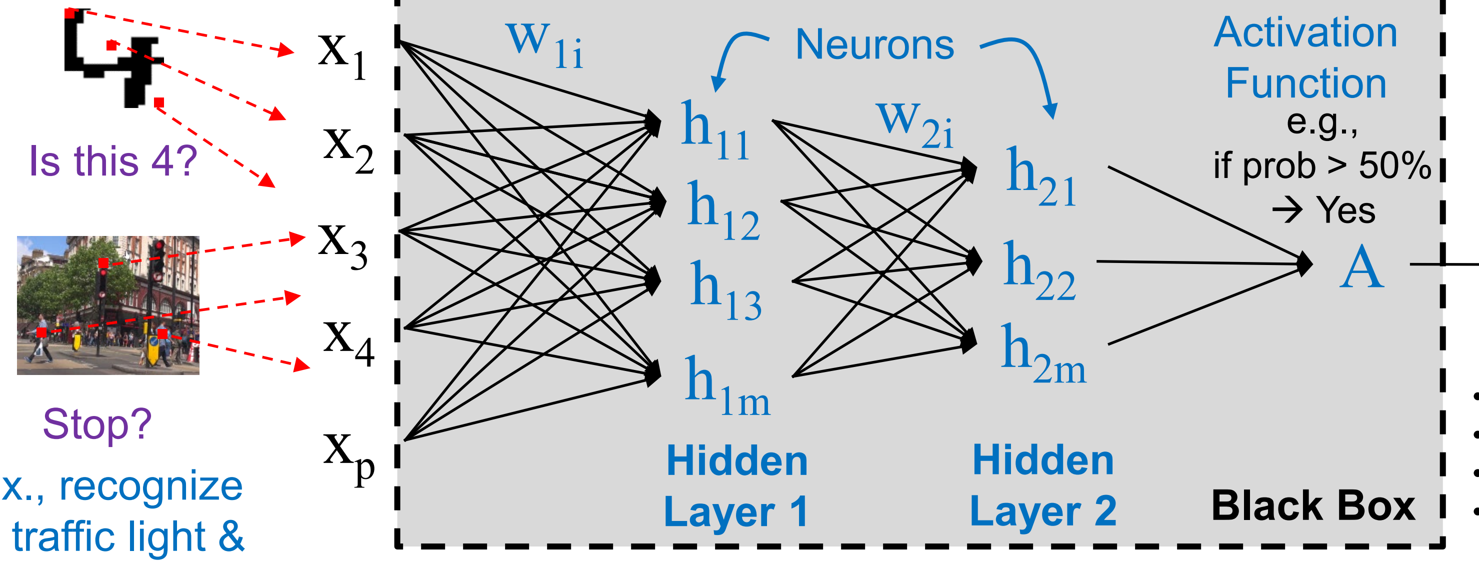
Artificial neural networks are built from nodes that are coded with a value. The nodes are connected to each other and, when the network is trained, the connections between nodes that are active at the same time get stronger, otherwise they get weaker.

©Johan Jarnestad/The Royal Swedish Academy of Sciences



ex. recognize hand-written zip code numbers

Input Layer



Is this 4?

Stop?

ex., recognize a traffic light & decide to stop or not

Yes

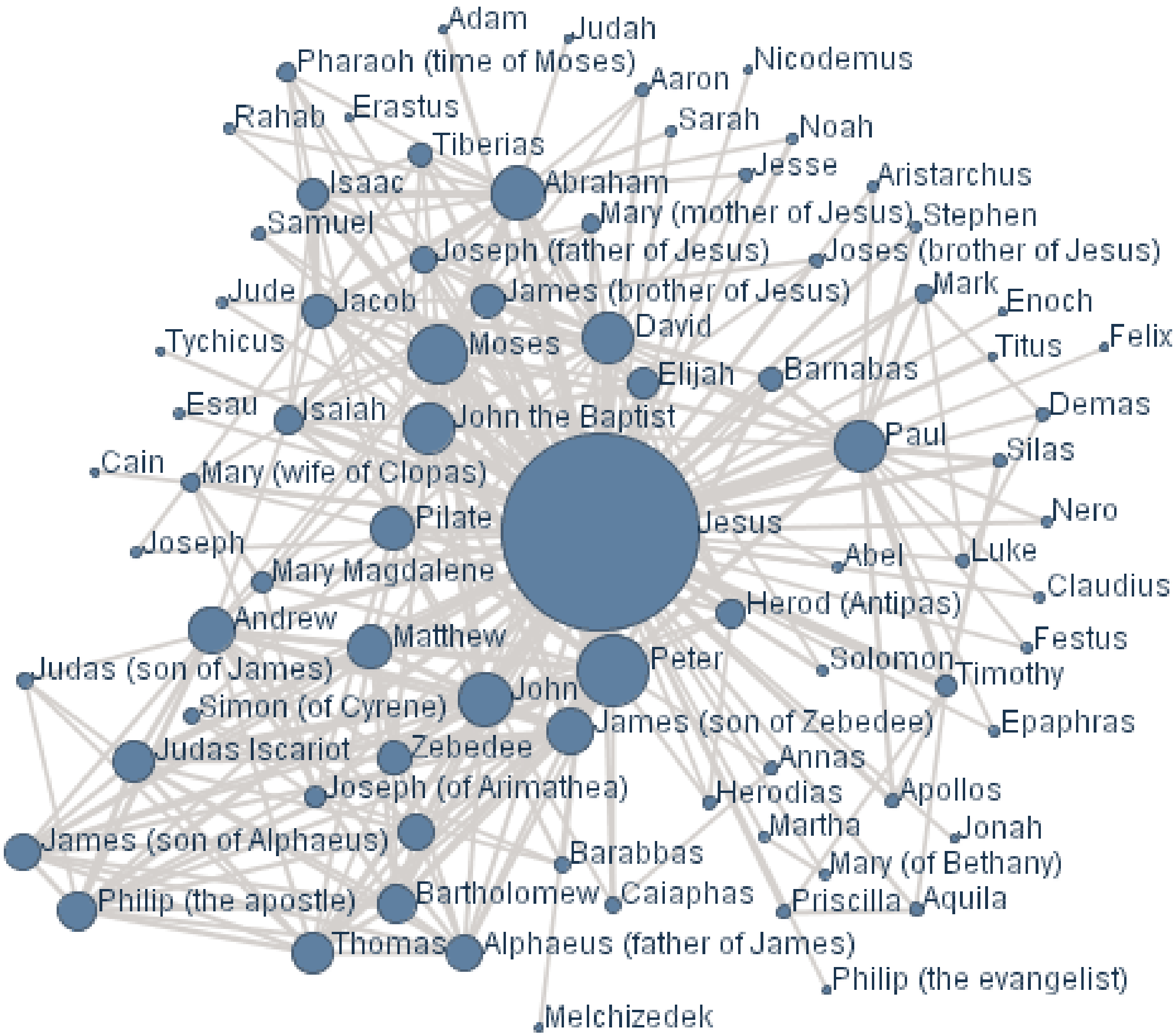
y

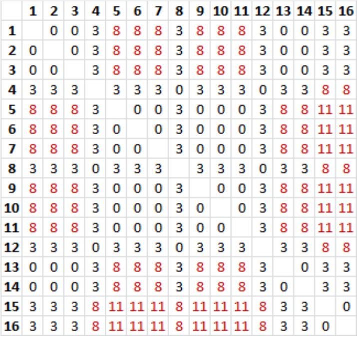
Outcome

- Values
- Classifications
- Words
- Images, etc.

Language Model Illustration: The Bible (by IBM)

Tie weight: number of times the pair appeared in the same passage.

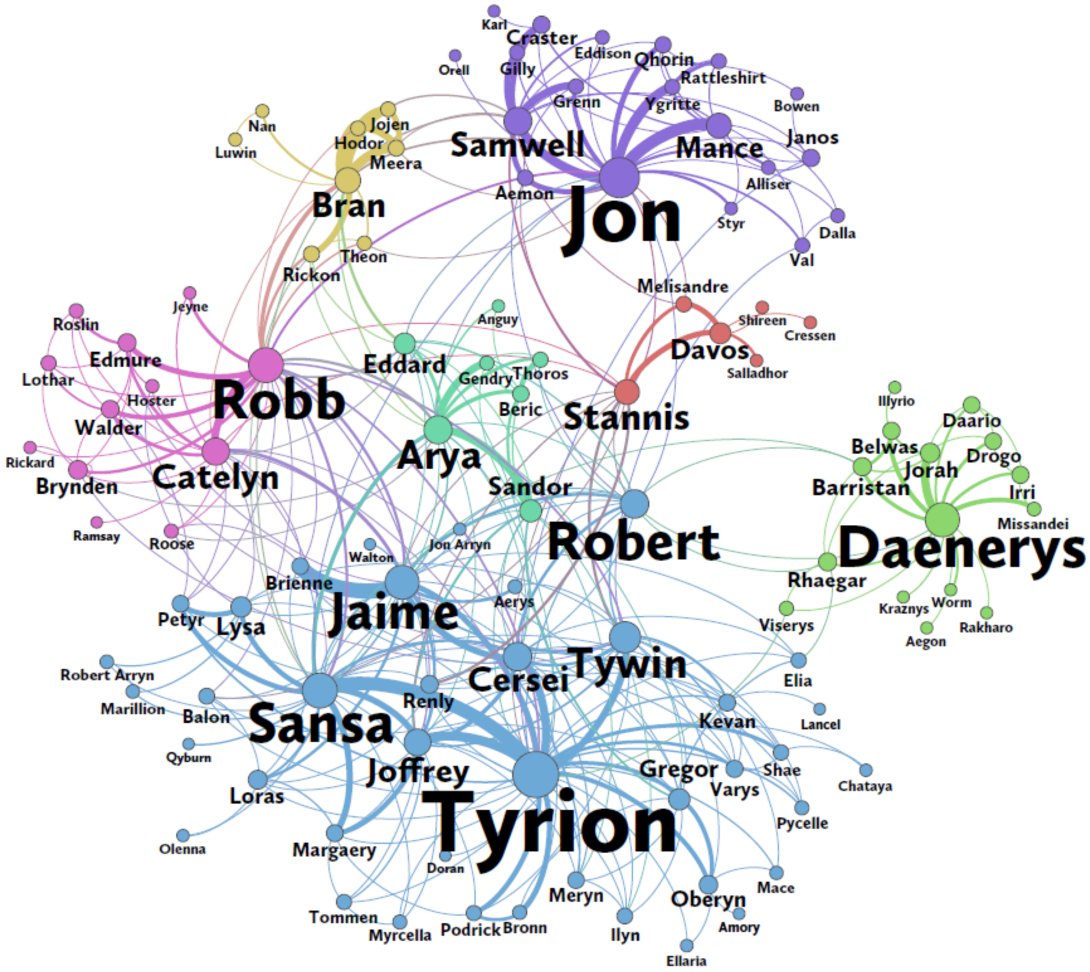


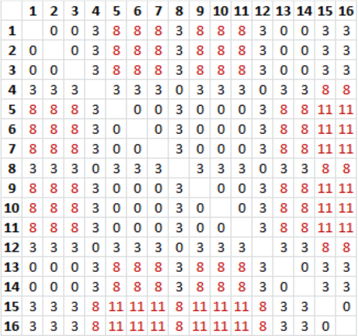


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0	0	3	8	8	8	3	8	8	8	3	0	0	3	3
2	0		0	3	8	8	8	3	8	8	8	3	0	0	3	3
3	0	0		3	8	8	8	3	8	8	8	3	0	0	3	3
4	3	3	3		3	3	3	0	3	3	3	0	3	3	8	8
5	8	8	8	3		0	0	3	0	0	0	3	8	8	11	11
6	8	8	8	3	0		0	3	0	0	0	3	8	8	11	11
7	8	8	8	3	0	0		3	0	0	0	3	8	8	11	11
8	3	3	3	0	3	3	3		3	3	3	0	3	3	8	8
9	8	8	8	3	0	0	0	3		0	0	3	8	8	11	11
10	8	8	8	3	0	0	0	3	0		0	3	8	8	11	11
11	8	8	8	3	0	0	0	3	0	0		3	8	8	11	11
12	3	3	3	0	3	3	3	0	3	3	3		3	3	8	8
13	0	0	0	3	8	8	8	3	8	8	8	3		0	3	3
14	0	0	0	3	8	8	8	3	8	8	8	3	0		3	3
15	3	3	3	8	11	11	11	8	11	11	11	8	3	3		0
16	3	3	3	8	11	11	11	8	11	11	11	8	3	3	0	

Illustration: Network of Thrones

Tie weighs: number of times the pair appeared within 15 words in the book



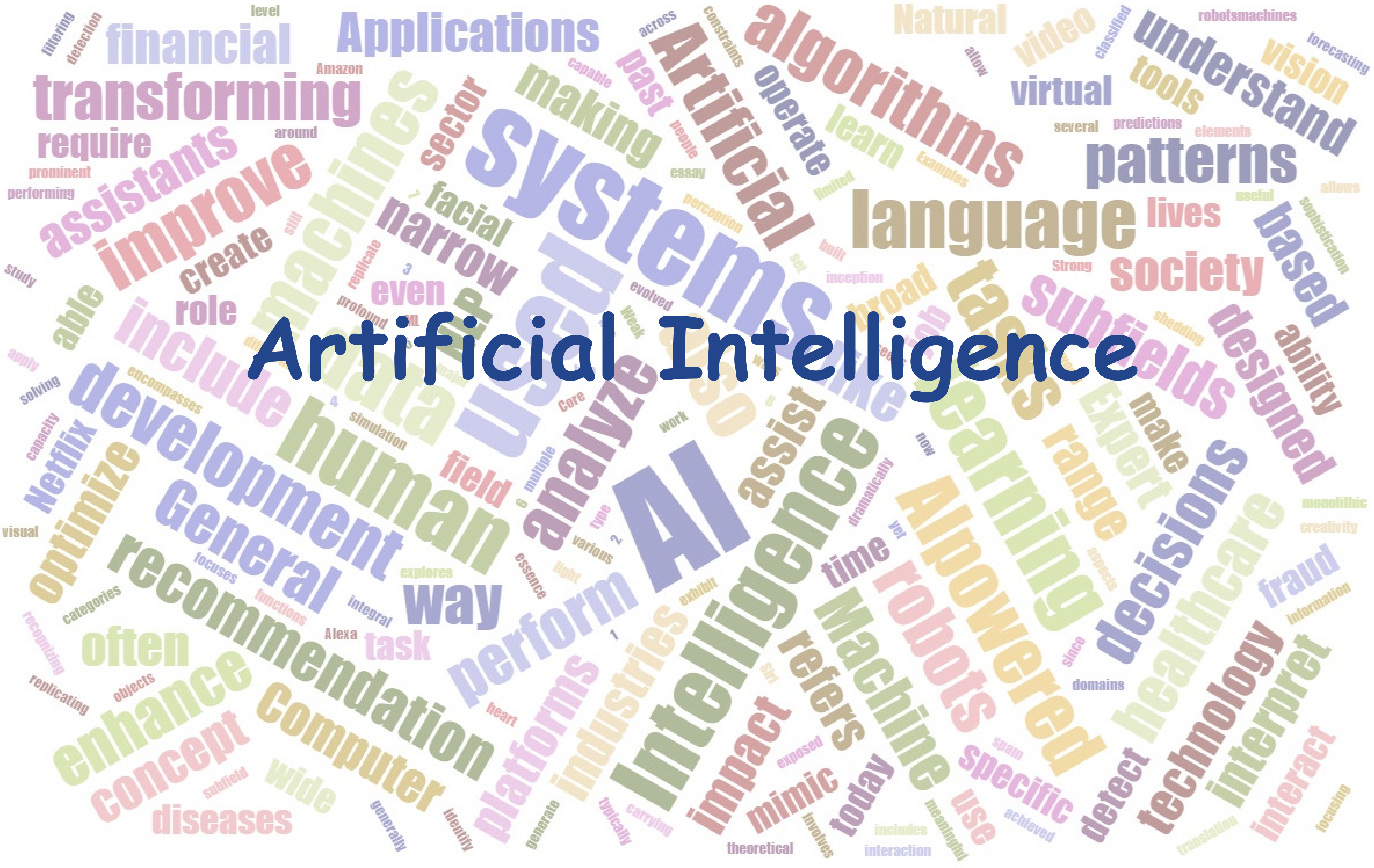


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0	0	3	8	8	8	3	8	8	8	3	0	0	3	3
2	0		0	3	8	8	8	3	8	8	8	3	0	0	3	3
3	0	0		3	8	8	8	3	8	8	8	3	0	0	3	3
4	3	3	3		3	3	3	0	3	3	3	0	3	3	8	8
5	8	8	8	3		0	0	3	0	0	0	3	8	8	11	11
6	8	8	8	3	0		0	3	0	0	0	3	8	8	11	11
7	8	8	8	3	0	0		3	0	0	0	3	8	8	11	11
8	3	3	3	0	3	3	3		3	3	3	0	3	3	8	8
9	8	8	8	3	0	0	0	3		0	0	3	8	8	11	11
10	8	8	8	3	0	0	0	3	0		0	3	8	8	11	11
11	8	8	8	3	0	0	0	3	0	0		3	8	8	11	11
12	3	3	3	0	3	3	3	0	3	3	3		3	3	8	8
13	0	0	0	3	8	8	8	3	8	8	8	3		0	3	3
14	0	0	0	3	8	8	8	3	8	8	8	3	0		3	3
15	3	3	3	8	11	11	11	8	11	11	11	8	3	3		0
16	3	3	3	8	11	11	11	8	11	11	11	8	3	3	0	

Source: https://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20%281%29.pdf

Current Trends in Machine Learning

- **Large Language Models:** generate human-like text, perform complex reasoning tasks, and even write code.
- **Generative AI:** respond to prompts and generate responses as text, images, sound, video, etc.
- **AutoML:** Automate model selection, feature engineering, and parameter tuning.
- **Foundation Models:** models like GPT that can be adapted to a wide range of applications.
- **Explainability for Complex Models:** Tools and methods are being developed to make models more transparent and to explain their predictions to non-experts.
- **Multimodal Models:** Can process and combine data from different sources (text, image, video, audio, etc.)



Artificial Intelligence

Artificial Intelligence (AI)

- **AI** will not replace jobs, but people who know AI will !!
- **AI** is an old discipline – started in 1956 (Darmouth Prof. John McCarthy)
- Defined as: *"the science and engineering of making computers behave emulating human intelligence"* → *learning, reasoning and self-correction.*
  - ➢ *Gartner → AI is the application of advanced analysis, logic-based methods and machine learning (ML) to interpret events and automate decisions.*
- There are many branches of AI, such as:
  - ✓ **Logic-Based** – based on logical rules
  - ✓ **Expert Systems** – AI "if-then" rules are elicited from experts
  - ✓ **Machine Learning (ML)** – *systems that learn mostly from historical data, without explicit programming*
  - ✓ **Generative AI (GAI)** – *"deep learning models that can generate high-quality text, images and other content based on data they were trained on"* (IBM) → generate content from user prompts.

Prompt Engineering

Is the process of designing, refining, and optimizing prompts when interacting with AI models like GPT (Generative Pre-trained Transformer). Key aspects of Prompt Engineering include:

- Clarity and Specificity → ensure the model understands what's asked
- Contextualization → provide general context to get accurate response
- Instructions for Tasks → state what's expected (e.g., summary)
- Iterative Testing → try different prompts and check output consistency

Prompt:	***ChatGPT Response:***
"I am a machine learning professor at a business school. How do I explain what prompt engineering is to an audience of accountants?"	*"To explain prompt engineering to an audience of accountants, you can draw parallels between how accountants optimize financial processes and how prompt engineers optimize AI outputs. Here's an example: Prompt engineering is the process of crafting precise and specific instructions for an AI model to ensure it provides accurate and relevant results, much like how you would design a financial query to get the correct data from accounting software. In the same way that vague or poorly structured inputs can lead to incorrect financial reports, unclear prompts can lead to less useful AI responses. By refining these prompts, we help the AI deliver exactly what we need, whether it's a financial summary, analysis, or any other task."*

Artificial Intelligence (AI) & Machine Learning Applications by Industry

- **Healthcare:** diagnostics, personalized treatment, robotic surgery, AI-driven diagnostic tools, drug discovery, predictive analytics for patient management.
- **Finance & Accounting:** risk management, fraud detection, accounting forensics, algorithmic trading, personalized banking services, and AI-driven compliance.
- **Retail:** personalized marketing, inventory management, sales forecasting, e-commerce, chatbots, recommendation engines.
- **Manufacturing:** production line optimization, predictive maintenance, quality control, automation.
- **Transportation:** self-driving cars, route optimization, logistics, traffic management, autonomous systems.
- **Energy:** energy use optimization, predicting equipment failure, managing grids, efficient energy consumption.
- **Advertising & Media:** content creation, targeted advertising, media analytics

A Few Useful General AI Tools

- **GhatGPT** (https://chatgpt.com/): conversational **G**enerative **P**re-Trained **T**ransformer AI model. GPT-3.5 is free, ChatGPT Plus is $20/month.
- **Dall-E** (https://openai.com/index/dall-e-3/): generates graphs in ChatGPT
- **MidJourney** (https://www.midjourney.com/home): Generates images from text.
- **Copilot** (https://www.microsoft.com/en-us/microsoft-copilot/personal-ai-assistant): Microsoft AI tool with free and paid versions.
- **Perplexity** (https://www.perplexity.ai/): Conversational AI, similar to ChatGPT, but very accurate and yields useful citations
- **Pearl.ai, AI Code Editor, Claude, Cursor:** generative AI tools with advanced features for software programming.
- **Notebook LM** (https://notebooklm.google.com/): By Google, for generative AI on local documents like Word, Powerpoint and PDF.
- **Poe** (https://poe.com/): Great to run generative AI over local documents like Word, PowerPoint, and PDF files and render useful summaries.
  See my Chatbot on Predictive Analytics: https://poe.com/ITEC-621
- **DeepL Translate** (https://www.deepl.com/en/translator): for accurate translation between multiple languages
- **Grammarly** (https://www.grammarly.com/): Excellent AI writing tool

Popular Uses of Generative AI

- Research → Search papers, summarize articles, find citations, etc.
- Writing → Write first drafts, edit writing, change styles (e.g., academic)
- Local Knowledge Base → Chat and generate from your own materials
- Local Coach/Tutor → Support learning and content creation
- Advanced Searches → Generate content, not just website links
- Create Presentations → Generate slides and graphics
- Generate Creative Content → Images, Music, Video, etc.
- Software Programming → Learn how to code, generate code, etc.
- Language Translation → Text or audio from/to any language
- Power Computer User → Tools (ex., Copilot for Excel, PowerPoint)
- Voice Applications → Voice assistants, transcriptions
- Face Recognition → Tagging photos, device security, etc.

Current Trends in Artificial Intelligence

- **Generative AI (GAI):** ChatGPT and other tools that respond to a prompt and generate content based on it. The content can be text, images, videos, software code, etc. resulting in innovation and more efficient work processes.
- **Augmented Intelligence:** **copilot** decision makers, doctors, manufacturers, artists, etc.
- **AI in Predictive Analytics:** fraud detection, cybersecurity analysis, healthcare predictions, trend and financial forecasting trends, decision making, resource allocation and optimization.
- **AI in Personalization and Marketing:** personalize customers' experience, product and service recommendation engines.
- **Responsible AI and Ethics:** creating transparent, explainable and unbiased AI systems.

Business Analytics @ Kogod

MS Analytics (being re-branded + AI)

http://fs2.american.edu/alberto/www/analytics/

Curriculum

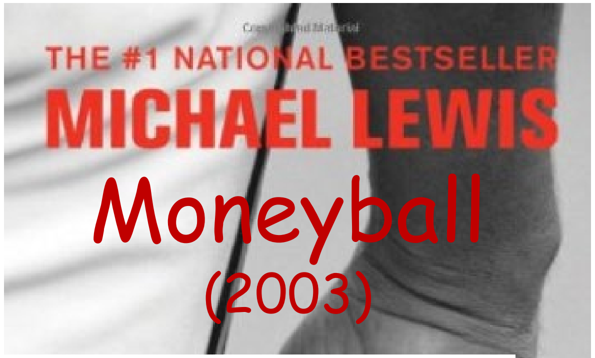
http://fs2.american.edu/alberto/www/analytics/msanalytics.pdf

Questions?

Backup Slides

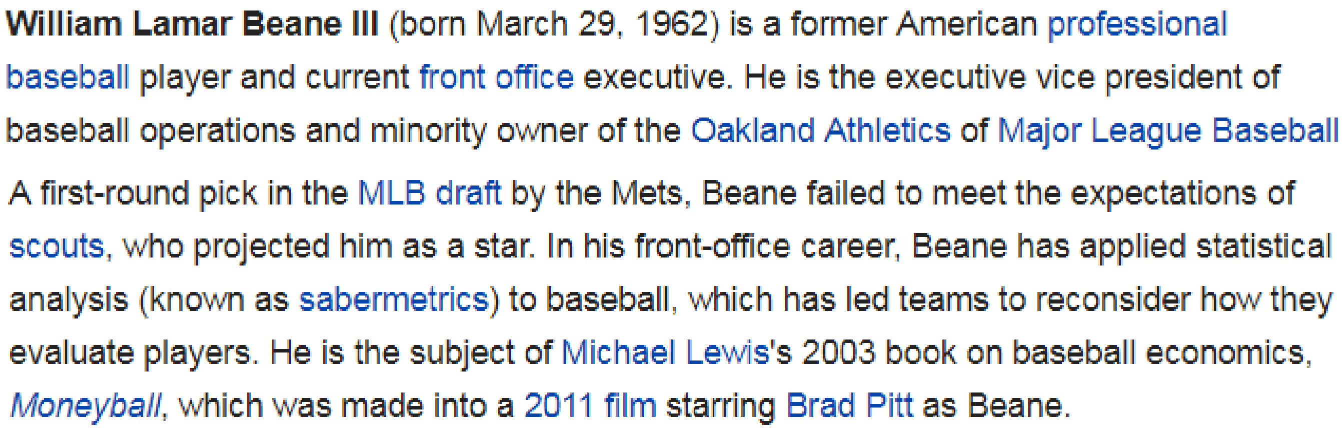
Moneyball book author: Michael Lewis

Billy Beane (Oakland A’s) incorporated (“**sabermetrics**”) analytics and statistics, reaching record profits in baseball



Moneyball
(2003)

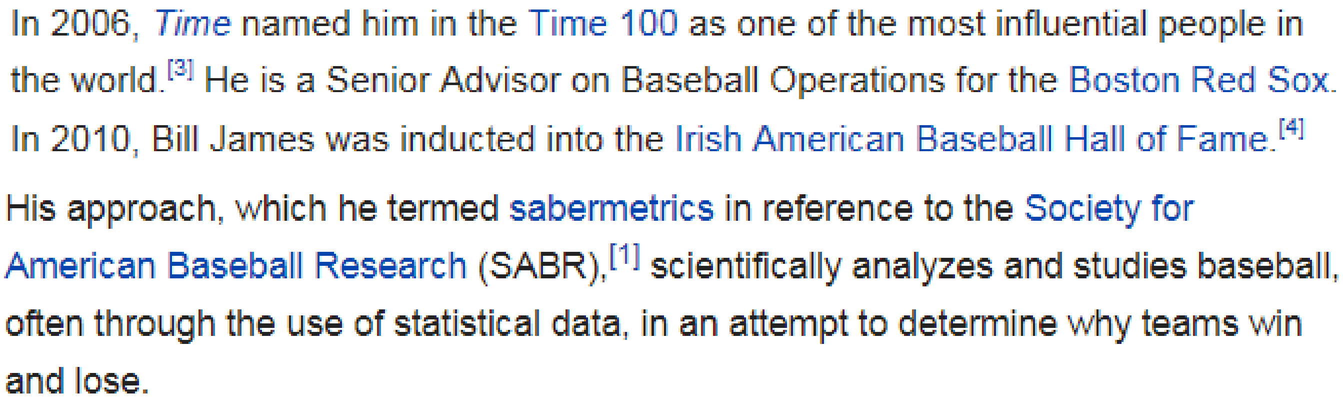

Billy Beane



William Lamar Beane III (born March 29, 1962) is a former American professional baseball player and current front office executive. He is the executive vice president of baseball operations and minority owner of the Oakland Athletics of Major League Baseball

A first-round pick in the MLB draft by the Mets, Beane failed to meet the expectations of scouts, who projected him as a star. In his front-office career, Beane has applied statistical analysis (known as sabermetrics) to baseball, which has led teams to reconsider how they evaluate players. He is the subject of Michael Lewis's 2003 book on baseball economics, *Moneyball*, which was made into a 2011 film starring Brad Pitt as Beane.

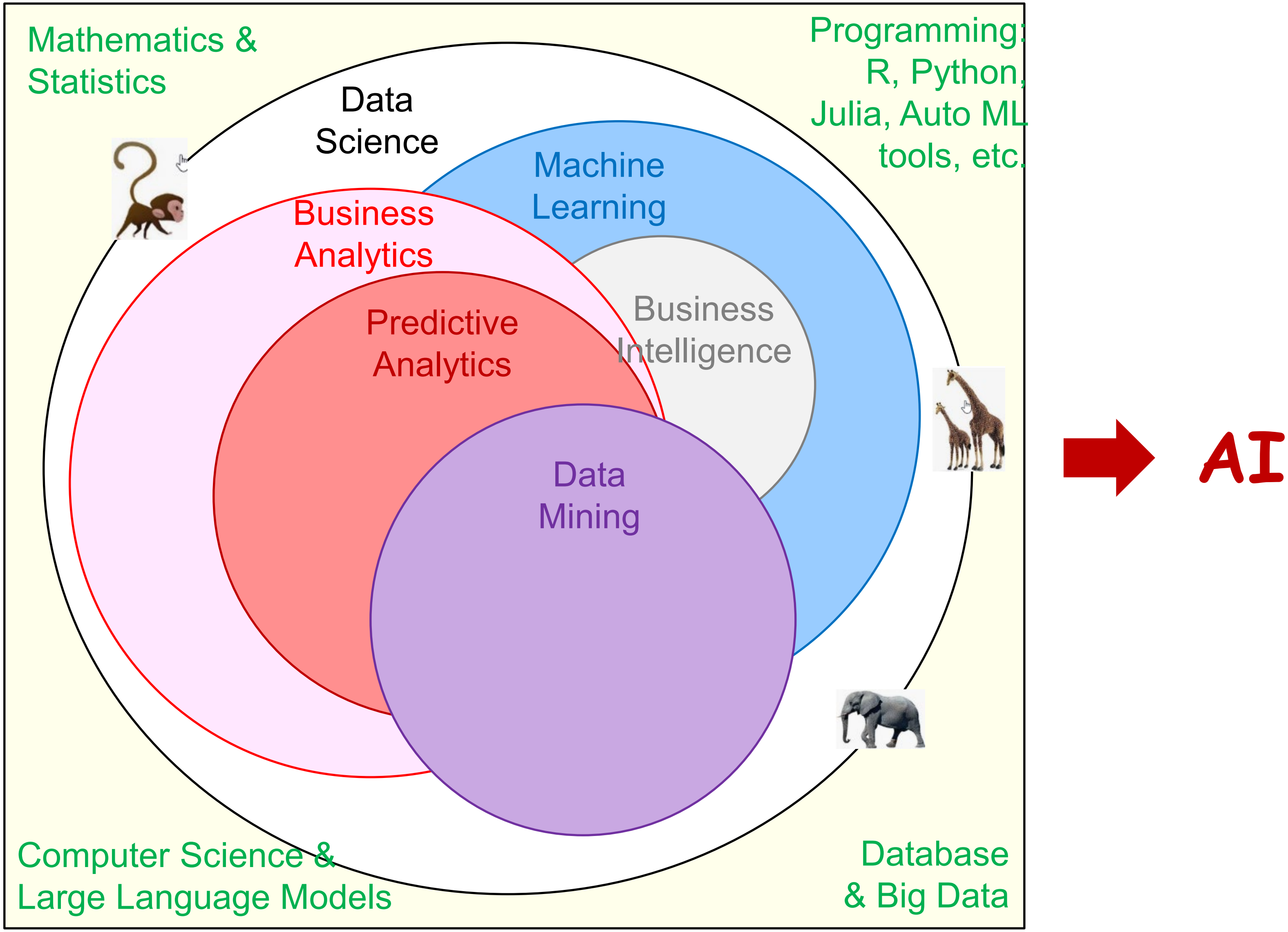
Bill James



In 2006, *Time* named him in the Time 100 as one of the most influential people in the world.[3] He is a Senior Advisor on Baseball Operations for the Boston Red Sox. In 2010, Bill James was inducted into the Irish American Baseball Hall of Fame.[4]

His approach, which he termed sabermetrics in reference to the Society for American Baseball Research (SABR),[1] scientifically analyzes and studies baseball, often through the use of statistical data, in an attempt to determine why teams win and lose.

AI & Data Science Zoo



Mathematics & Statistics

Programming: R, Python, Julia, Auto ML tools, etc.

Data Science

Machine Learning

Business Analytics

Predictive Analytics

Business Intelligence

Data Mining

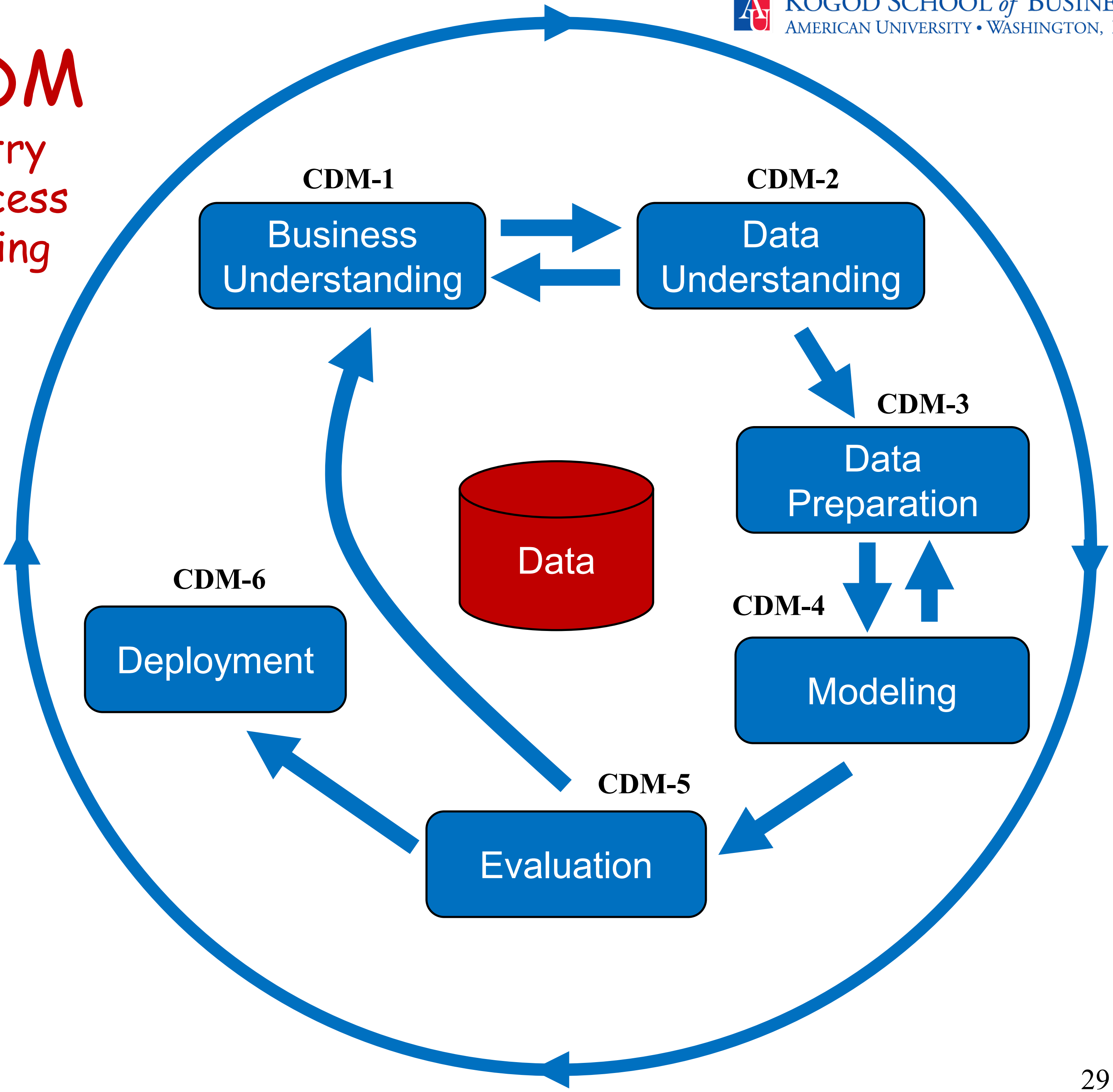
Computer Science & Large Language Models

Database & Big Data

→ AI

CRISP-DM

Cross-Industry Standard Process for Data Mining



Main Components of Analytics

Insights

- Report
- Meaning & Interpretation
- Storytelling

Models

- Model Specification
- Model Method
- Interpretable vs. Black Box
- Model Assumptions
- Large Language Models
- Machine Learning
- Artificial Intelligence

Data

- Structured (Tables)
- Unstructured
- Big Data
- Text
- Data Pre-Processing & ETL

Business Question / Problem / Case

- Quantitative Outcome
- Classification Outcome

Kogod MS Analytics

KSB 620 & 621 Capstone Experience Practicum (2 x 1.5 cr)

Functional Specialization (12 cr)

- Business Administration
- Machine Learning & AI
- Marketing
- IT&A Consulting
- Data Science
- Accounting Forensics
- Accounting Analytics
- Financial Quantitative Analysis
- Sustainability Management

Analytics Core (9 cr)

- ITEC 610 Managerial Statistics
- ITEC 620 Business Insights Through Analytics
- ITEC 621 Predictive Analytics

IT Core (9 cr)

- ITEC 616 Management Information Systems
- ITEC 670 Database & Big Data
- ITEC 660 Business Intelligence
- Tools:
  KSB 999 R;
  ITEC 600 Python

The Analytics Life Cycle

Step 1 – Business Understanding (CDM - 1)

(1) Formulate Business **Problem** and **Case**

(2) Translate into **Analytics Question**: outcome + predictors

Step 2 – Data Work (CDM - 2, 3)

Identify & Gather data: structured, unstructured, visual, etc.

Pre-process data: cleanse, prepare, transform, format, etc.

Descriptive Analytics: e.g., correlation, descriptive stats, etc.

Step 3 – Modeling Method (CDM - 4)

Predictive Analytics → predict outcomes

- **Goals**: Interpretability vs. Accuracy
- **Model Type:** Explainable (e.g., regression) vs. Black Box (e.g., neural network)
- **Prediction Type:** Quantitative vs. Classification Outcome
- **Model Evaluation:** Model assumptions, fit statistics, cross validation testing
- **Model Specification:** Predictors or features? Linear, non-linear, log?

Step 4 – Analysis (CDM - 5)

Interpret **effects**, test hypotheses and/or make predictions

Step 5 – Reporting (CDM - 6)

Written, interactive, visual → **"storytelling"**

Model Explainability

Interpretable Models

Sales ~ Ad Spending + Store Size

Explain how the model works and what the results **mean**
→ e.g., OLS regression, logistic regression, decision trees
→ good for **interpretability** and **inference**

Black Box Models (non-interpretable)

The complexities of the internal model are **difficult** to explain
→ e.g., neural networks, deep learning
→ good for **predictive accuracy**

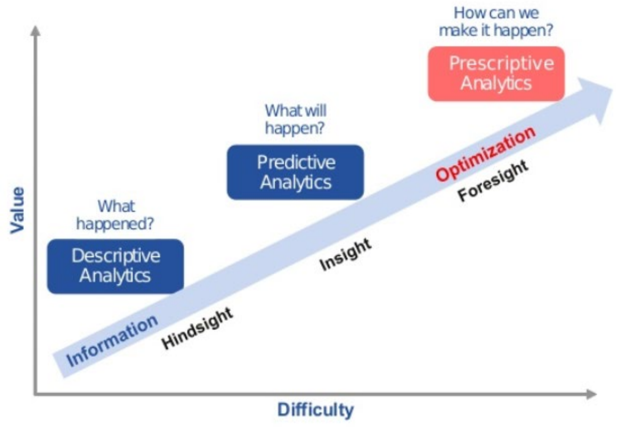
Stop?

Analytics Types

- **Descriptive:** getting meaning from the data – e.g., descriptive statistics, data mining, cluster analysis, market basket analysis, etc.

- **Predictive:** using some data to predict outcomes (quantitative or classification) – e.g., projecting sales/profits, decease survival prognosis, probability of losing/gaining a client, probability of a security breach, etc.

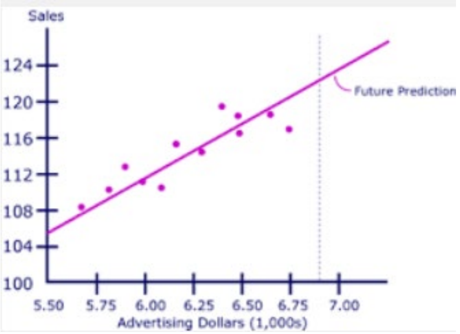
- **Prescriptive:** decision models that use predictive models to make/automate recommendations – e.g., optimal pricing, where to drill for oil, selecting job applicants, market products based on purchase history, etc.



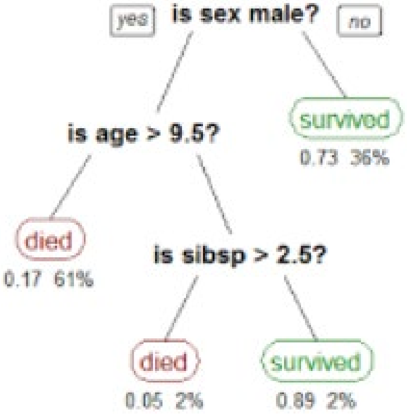
Derived from: Gartner (December 2012)

Predictive Analytics Question → Outcome → What are you predicting?

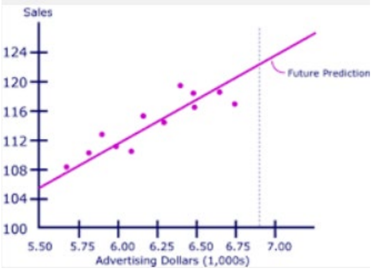
- **Values:** → predict **quantitative** outcomes – e.g., how much does annual income increase with each year of additional university education; how much sales increase for each additional dollar of advertising expenditures?



- **Classification:** predict **categorical** outcomes or the likelihood that an observation will fall in a particular class – e.g., yes/no, positive/negative illness diagnosis, loan default/no-default, cyber security breach/no-breach, etc.

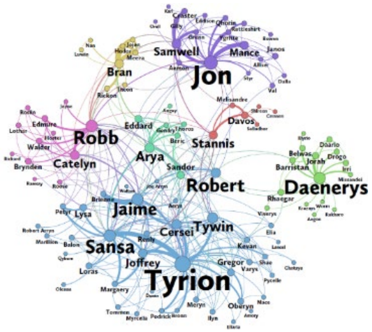


Analytics Approaches



- **Quantitative:** statistical, association, correlation, regression, machine learning, etc. models based on how variables co-vary with the outcome – e.g., how much does annual income increase with each year of additional university education?

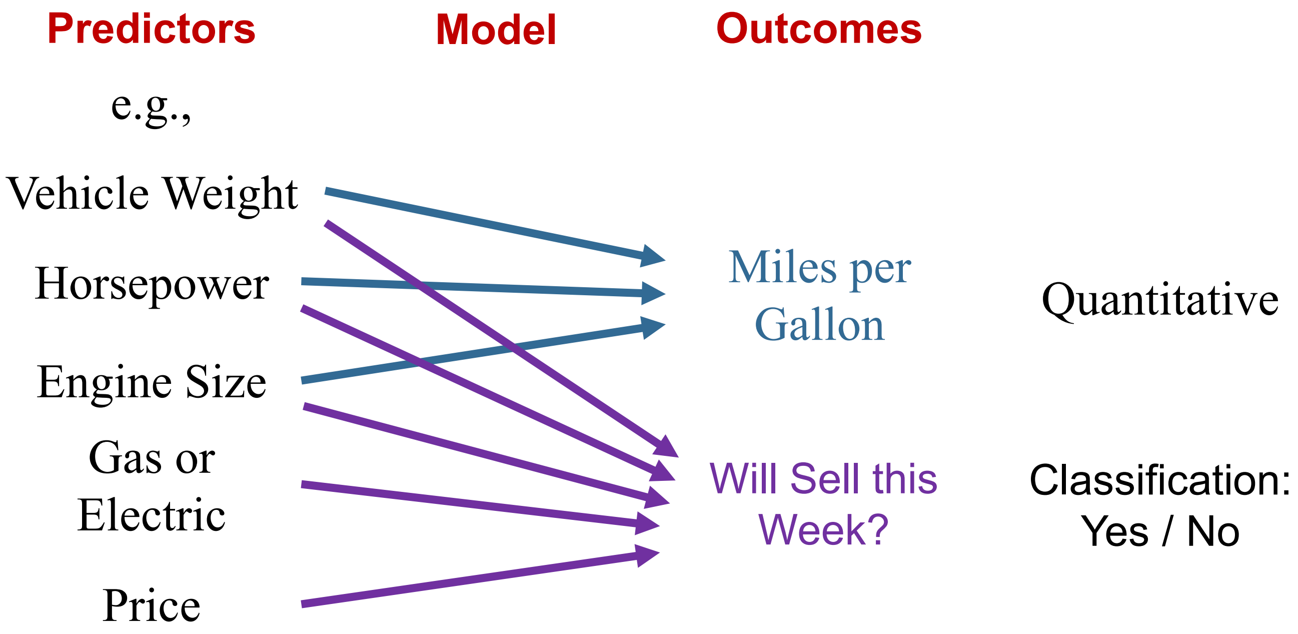
- **Visual:** e.g., Tableau, R, Python



- **Network Analytics:** relationships (e.g., friendship, communication, co-memberships, etc.) → social, organizational, professional networks, etc.
- **Unstructured Data Analysis:** text mining, natural language processing, social media, etc.
- **Large Language Models (LLMs):** predict the next word based on the current set of words in the text → generate text responses, e.g., ChatGPT

Statistical Models

- **Regression** → predicting by statistical association using variables in the data to predict quantitative outcomes or classifications.
- **Tree** → partitioning the data at specific values of variables to differentiate quantitative outcomes or classifications



Predictors **Model** **Outcomes**

e.g.,

Vehicle Weight

Horsepower

Engine Size

Gas or Electric

Price

Miles per Gallon — Quantitative

Will Sell this Week? — Classification: Yes / No

ics Modeling Options

- Ordinary Least Squares
- Weighted Least Squares
- Generalized Linear Model
- Ridge Regression
- LASSO Regression
- Principal Components Regression
- Partial Least Squares
- Non-Linear Models
- Piecewise Linear Regression
- Step Regression
- Linear, Cubic, Smoothing Splines
- Neural Networks
- Etc.

	Modeling Method		
	Structured		Visual, Text, uctured, etc.
	analysis, correlation, ma ysis, sample statistics,		charts, network rams, natural ge processing, ng dendograms, etc.
Predictive	**Association**	**Decision Tree**	**Charts**
Quantitative Value	Regression	Regression Trees	Regression plots, scatter plots, Tableau diagrams, trend charts, etc.
Classification	Logistic Regression; Other Categorical Regression Models		Tree maps, interactive agrams,
Prescriptive	Operations research, optimization, linea		lations, etc.

- Regression Trees
- Classification Trees
- Bootstrap Aggregation
- Random Forest
- Boosting Trees
- Etc.

- Binomial Logistic Regression
- Multinomial Logistic Regression
- Linear Discriminant Analysis
- Quadratic Discriminant Analysis
- Neural Networks
- Etc.

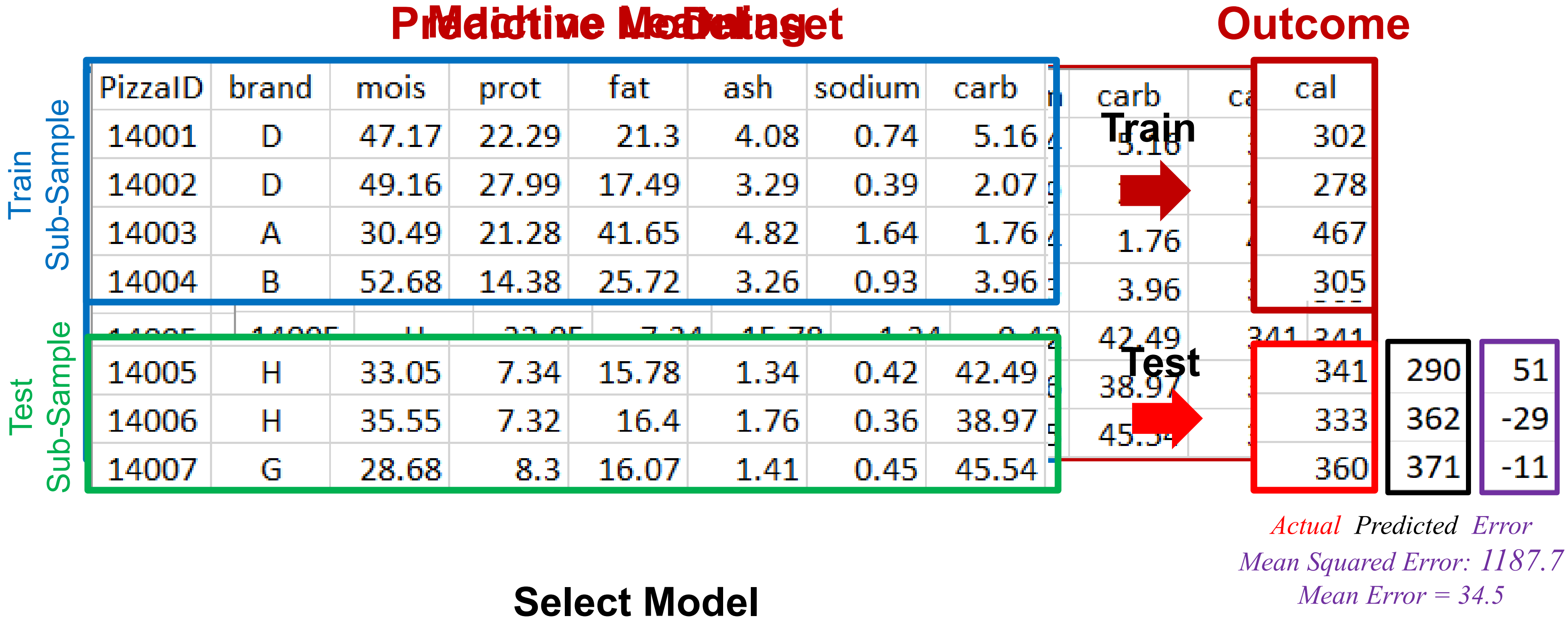
ML: Key Concepts

- Again, **ML** is about **maximizing predictive accuracy** →

1. **Split** the data set into **train** and **test** sub-samples
2. **Train subset** → part of the data selected randomly to **fit models** and estimate parameters (e.g., regression coefficients)
3. **Test subset** → part of the data, **not** in the **train** subset, used to **evaluate** the trained **models**
4. **Cross-Validation (CV)** → Evaluate the trained model’s predictive accuracy with the test subset, → CV Test Error
5. **Re-Sample** and repeat 1-5 many times

Machine Learning

Cross-Validation → Split, Train, Test, Re-Sample



Predictive Model / **Machine Dataset**

Outcome

Train Sub-Sample

PizzaID	brand	mois	prot	fat	ash	sodium	carb	cal
14001	D	47.17	22.29	21.3	4.08	0.74	5.16	302
14002	D	49.16	27.99	17.49	3.29	0.39	2.07	278
14003	A	30.49	21.28	41.65	4.82	1.64	1.76	467
14004	B	52.68	14.38	25.72	3.26	0.93	3.96	305

Train

Test Sub-Sample

PizzaID	brand	mois	prot	fat	ash	sodium	carb	Actual	Predicted	Error
14005	H	33.05	7.34	15.78	1.34	0.42	42.49	341	290	51
14006	H	35.55	7.32	16.4	1.76	0.36	38.97	333	362	-29
14007	G	28.68	8.3	16.07	1.41	0.45	45.54	360	371	-11

Test

Actual *Predicted* *Error*

Mean Squared Error: 1187.7

Mean Error = 34.5

Select Model

New Data

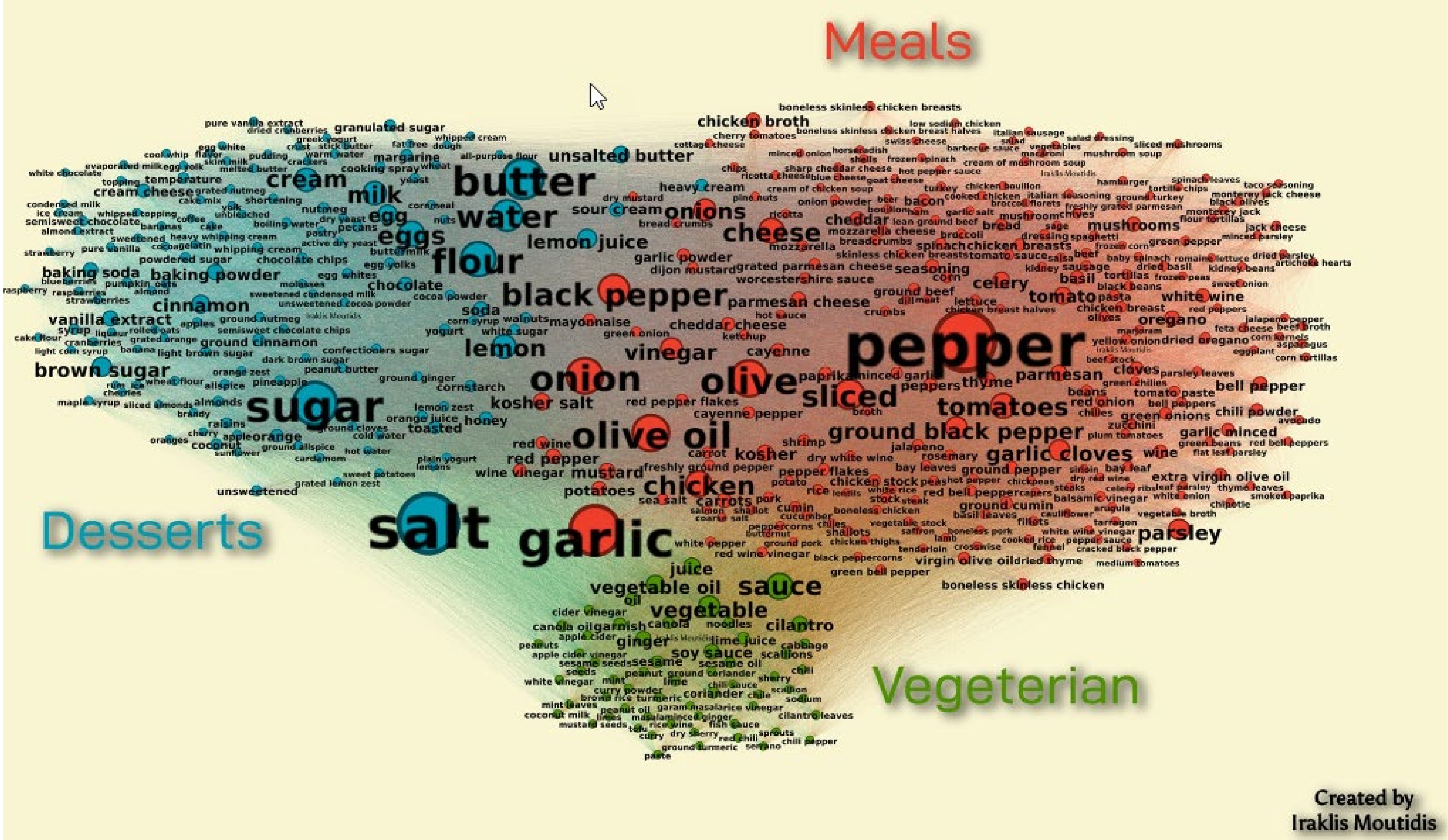
D	49.16	27.99	17.49	3.29	0.39	2.07

→

Predict:
- *Values*
- *Classifications*
- *Words*
- *Images, etc.*

Uses: test and compare models, tune model parameters, train deep and supervise learning models, etc.

Illustration: Food Ingredient Network



Created by
Iraklis Moutidis